LECTURES ON FUNDAMENTAL CONCEPTS OF ALGEBRA AND GEOMETRY

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO
SAN FRANCISCO

MACMILLAN & CO., LIMITED
LONDON · BOMBAY · CALCUTTA
MELBOURNE

THE MACMILLAN CO. OF CANADA, LTD.
TORONTO

LECTURES

ON

FUNDAMENTAL CONCEPTS OF ALGEBRA AND GEOMETRY

BY

JOHN WESLEY YOUNG

PROFESSOR OF MATHEMATICS IN THE UNIVERSITY OF KANSAS

PREPARED FOR PUBLICATION WITH THE COOPERATION OF

WILLIAM WELLS DENTON

ASSISTANT IN MATHEMATICS IN THE UNIVERSITY OF ILLINOIS

WITH A NOTE ON

THE GROWTH OF ALGEBRAIC SYMBOLISM

BY

ULYSSES GRANT MITCHELL

ASSISTANT PROFESSOR OF MATHEMATICS IN THE UNIVERSITY OF KANSAS

New York

THE MACMILLAN COMPANY

1923

Set up and electrotyped. Published June, 1911.

Norwood Press
J. S. Cushing Co. — Berwick & Smith Co.
Norwood, Mass., U.S.A.

PREFACE

THE following lectures contain an elementary account of the logical foundations of algebra and geometry,— elementary, in the sense that the technical mathematical equipment presupposed on the part of the reader has been reduced to a minimum. Except in a very few instances, no knowledge of mathematics beyond the most elementary portions of algebra and geometry has been assumed. It has been my purpose to give a general exposition of the abstract, formal point of view developed during the last few decades, rather than an exhaustive treatment of the details of the investigations.

The results of recent work on the logical foundations are of vital interest alike to the teachers of mathematics in our secondary schools and colleges and to philosophers and logicians. I hope that both these classes will welcome a concise statement of some of the more fundamental of these results and an elementary exposition, omitting all involved details, of the point of view which governs all present work on the foundations. The book should be available also as a text in connection with so-called Teachers' Courses in colleges and universities.

The lectures were given at the University of Illinois during the summer of 1909. They are here reproduced in

substantially the same form as delivered. The conversational style has, to a large extent, been retained in the hope that the presentation has gained thereby in spontaneity.

My cordial thanks are due my former colleagues, Dean E. J. Townsend and Professor G. A. Miller, of the University of Illinois, who read the greater part of the manuscript; and to my colleague, Professor U. G. Mitchell, who not only read the whole manuscript and rendered valuable assistance in seeing the book through the press, but has added to its value by contributing the Note on "The Growth of Algebraic Symbolism," which will be found at the end of the lectures. Above all, however, my thanks are due to Mr. W. W. Denton, of the University of Illinois, without whose help the lectures would probably not have been published. He took the lectures down stenographically, and applied himself to the revision of the resulting manuscript with great enthusiasm and keen insight.

J. W. YOUNG.

LAWRENCE, KANSAS,
April, 1911.

CONTENTS

FUNDAMENTAL CONCEPTS OF ALGEBRA AND GEOMETRY

LECTURE I

INTRODUCTION. EUCLID'S ELEMENTS

Two Aspects of Mathematics.—Mathematics may be considered from two aspects. The first, or utilitarian, regards mathematics as presenting in serviceable form a body of useful information. The second and educationally more important aspect, the one with which we shall chiefly concern ourselves in these lectures, relates to the fact that mathematics, in particular algebra and geometry, consists of a body of propositions that are *logically connected.* It is proposed to consider the more important fundamental concepts of algebra and geometry with regard to their logical significance and their logical interrelations. Let it be said at the outset that we shall not be primarily concerned with the psychological genesis of these concepts, nor with the manifold and interesting philosophical questions to which they give rise.

"Mathematical Science" defined.—We are at once confronted with the question: What is mathematics? To give a satisfactory definition is difficult, if not impossible. We

shall be in a better position to appreciate the difficulties attaching to this question at the close of the lectures. We may, however, define what we shall understand by a mathematical science. A *mathematical science*, as we shall use the term, is any body of propositions arranged according to a sequence of logical deductions; *i.e.* arranged so that every proposition of the set *after a certain one* is a formal logical consequence of some or all of the propositions that precede it.[1] This definition is open to the criticism that it is too broad; it contains more than is usually understood by the term it professes to define. The idea, however, is simply that whenever a body of propositions is arranged or can be arranged in a strictly logical sequence, then by virtue of that fact we may call it mathematical. It will do no harm, if the meaning we attribute to this term in the present connection is broader than that usually attributed to it; the considerations that follow merely have a wider field of application.

Unproved Propositions and Undefined Terms. — Let us suppose that we have before us a body of propositions satisfying this definition, and let us inquire what it must have for a point of departure. The first proposition cannot, of course, be a logical consequence of a preceding proposition of the set. The second, if the body of propositions is at all extensive, is probably not deducible from the first; for the logical implications of a single proposition are not many.

[1] This definition is closely related to a definition given by BENJAMIN PEIRCE, when he said that "mathematics is the science which draws necessary conclusions."

If we consider the nature of a deductive proof, we recognize at once that there must be a hypothesis. It is clear, then, that *the starting point of any mathematical science must be a set of one or more propositions which remain entirely unproved.* This is essential; without it a vicious circle is unavoidable.

Similarly we may see that there must be some *undefined terms.* In order to define a term we must define it in terms of some other term or terms, the meaning of which is assumed known. In order to be strictly logical, therefore, a set of one or more terms must be left entirely undefined. One of the questions to be considered relates to the logical significance of the undefined terms and the unproved propositions. The latter are usually called axioms or postulates. Are these to be regarded as self-evident truths? Are they imposed on our minds *a priori,* and is it impossible to think logically without granting them? Or are they of experimental origin? Are the undefined terms primitive notions, the meaning of which is perfectly clear without definition? Closely connected with these questions are others relating to the validity of the propositions derived from the unproved propositions involving these undefined terms. We often hear the opinion expressed that a mathematical proposition is certain beyond any possibility of doubt by a reasonable being. Will a critical inspection bear out this opinion? We shall soon see that it will not. As an illustration of an extreme view, we may cite a definition of mathematics recently given by Bertrand Russell, one of the most eminent mathematical logicians of the present time. "Mathematics," he said, "is the science in which we never know

what we are talking about, nor whether what we say is true."[1] It is probable that many of our pupils will heartily concur in this definition. We shall see later that there is a sense in which this more or less humorous dictum of Russell is correct.

The Teaching of Mathematics. — There should be no need of emphasizing the importance of the questions just referred to. They lie at the basis of all science; every one interested in the logical side of scientific development is vitally concerned with them. Moreover, the general educated public shows signs of interest. The articles appearing from time to time in our popular magazines on the subjects of non-euclidean geometry and four-dimensional space give evidence hereof. It is merely one of the manifestations of the awakened popular interest in scientific progress.

These questions are, however, of particular interest to teachers of mathematics in our schools and colleges. Whether we regard mathematics from the utilitarian point of view, according to which the pupil is to gain facility in using a powerful tool, or from the purely logical aspect, according to which he is to gain the power of logical inference, it is clear that the chief end of mathematical study must be to make the pupil *think*. If mathematical teaching fails to do this, it fails altogether. The mere memorizing of a demonstration in geometry has about the same educational value as the memorizing of a page from the city directory. And yet it must be

[1] B. Russell, "Recent Work on the Principles of Mathematics," *The International Monthly*, vol. 4 (1901), p. 84.

admitted that a very large number of our pupils do study mathematics in just this way. There can be no doubt that the fault lies with the teaching. This does not necessarily mean that the fault is with the individual teacher, however. Mathematical instruction, in this as well as in other countries, is laboring under a burden of century-old tradition. Especially is this so with reference to the teaching of geometry. Our texts in this subject are still patterned more or less closely after the model of EUCLID, who wrote over two thousand years ago, and whose text, moreover, was not intended for the use of boys and girls, but for mature men.

The trouble in brief is that the authors of practically all of our current textbooks lay all the emphasis on the formal logical side, to the almost complete exclusion of the psychological, which latter is without doubt far more important at the beginning of a first course in algebra or geometry. They fail to recognize the fact that the pupil has reasoned, and reasoned accurately, on a variety of subjects before he takes up the subject of mathematics, though this reasoning has not perhaps been formal. In order to induce a pupil to think about geometry, it is necessary first to arouse his interest and then to let him think about the subject *in his own way* This first and difficult step once taken, it should be a comparatively easy matter gradually to mold his method of reasoning into a more formal type. The textbook which takes due account of this psychological element is apparently still unwritten, and as the teacher is to a large extent governed

by the text he uses, the failure of mathematical teaching is not altogether the fault of the teacher.

The latter must be prepared, however, to make the best of existing conditions. Much can be accomplished, even with a pedagogically inadequate text, if the teacher succeeds in awakening and holding his pupils' interest. It is well known that interest is contagious. Let the teacher be vitally, enthusiastically interested in what he is teaching, and it will be a dull pupil who does not catch the infection. It is hoped that these lectures may tend to give a new impetus to the enthusiasm of those teachers who have not as yet seriously considered the logical foundations of mathematics. Every thoughtful teacher has doubtless been confronted with certain logical difficulties in the treatment of topics in algebra and geometry. Even on the assumption that he has not had the hardihood of questioning the axioms and postulates which he finds placed at the basis of his science, — and it is hardly to be expected that he should thus question the validity of propositions which stood unchallenged for over two thousand years, — many serious difficulties attach to such topics as irrational numbers and ratios, complex numbers, limits, the notion of infinity, etc. How serious some of these difficulties are is made evident by the fact that in spite of the attention they received during several centuries, a satisfactory treatment has been found only within the last hundred years. Indeed, the present abstract point of view, which is to be described in these lectures, has been developed only within the last three or four decades.

Historical Development to be emphasized. — It is proposed throughout to emphasize the historical development of the conceptions and points of view considered. It is hoped hereby to give a comprehensive view of mathematical progress in so far as it relates to fundamental principles. This should tend to eradicate the all too common feeling that the fundamental conceptions of mathematics are fixed and unalterable for all time. Quite the contrary is the case. Mathematics is growing at the bottom as well as at the top; indeed, not the least remarkable results of mathematical investigation of recent years and of the present time relate to the foundations. Let the teacher once fully realize that his science, even in its most elementary portions, is alive and growing, let him take note of the manifold changes in point of view and the new and unexpected relations which these changes disclose, let him further take an active interest in the new developments, and indeed react independently on the conceptions involved, — for an enormous amount of work still remains to be done in adapting the results of these developments to the requirements of elementary instruction, — let him do these things, and he will bring to his daily teaching a new enthusiasm which will greatly enhance the pleasure of his labors and prove an inspiration to his pupils.

Results not of Direct Use in Teaching. — Reference has just been made to the need of *adapting* the results of the recent work on fundamental principles to the needs of the classroom. It should here be emphasized, perhaps, that the points of view to be developed in these lectures and the

results reached are not directly of use in elementary teaching. They are extremely abstract, and will be of interest only to mature minds. They should serve to clarify the teacher's ideas, and thus indirectly to clarify the pupil's. The latter's ideas will, however, differ considerably from the former's. The results referred to do, nevertheless, have a direct bearing on some of the pedagogical problems confronting the teacher. This will be discussed briefly as occasion arises.

Euclid's Elements of Geometry. — We propose in the first five lectures to consider rather informally our conceptions of space, and to illustrate in a general way the point of view to be followed in the later, more formal discussion. True to our purpose of taking into account the historical development of the conceptions involved, we can do no better than consider briefly at this point the fundamental notions that are found at the beginning of the earliest work in which mathematics is exhibited as a logically arranged sequence of propositions. I refer, of course, to EUCLID'S *Elements of Geometry*. This is the first attempt of which we have any record to establish a mathematical science as we have defined the term. Euclid lived about the year 300 B.C., and his greatest claim to fame is the fact that he furnished the succeeding centuries with the ideal of such a mathematical science. There is no doubt that it was his purpose to derive the properties of space from explicitly stated definitions, axioms, and postulates, without the use of any further assumptions, in particular without any further appeal to geometric intuition. It is true that he

made use of many propositions which he did not prove and which he did not explicitly state as unproved. But there is much evidence to show that his ideal was in accordance with our definition of a mathematical science. We may use Euclid's Elements as a convenient starting point to introduce the order of ideas which is to engage our attention. Any attempt to criticize Euclid's treatment of geometry is rendered peculiarly difficult at the outset on account of the great uncertainty that exists as to the real content of Euclid's text. Although he lived, as has been stated, about the year 300 B.C., the oldest manuscripts which purport to give Euclid's Elements date from about the year 900 A.D.[1] An interval of twelve hundred years intervenes between the time at which Euclid wrote and any record we have of his work. Moreover, there are several manuscripts dating from that time, and they differ considerably from one another.

Definitions. — How, then, did Euclid begin his treatment of geometry? We have seen what the starting point ought to be. It ought to be *a set of undefined terms and a set of unproved propositions such that every other term can be defined in terms of the former and every other proposition derived from the latter by the methods of formal logic* Euclid does indeed begin with a series of definitions, of which we will give a few examples:

A *point* is that which has no parts.

A *line* is length without breadth.[2]

[1] KLEIN, *Elementarmathematik vom hoheren Standpunkte aus*, vol II, p 404

[2] Some MSS. add: The extremities of a line are points.

A *straight line* is a line which lies evenly between two of its points.

These definitions serve to illustrate how it is necessary to define a term in terms of something else, the meaning of which is assumed known. The terms "part," "length," "breadth," "lies evenly" are undefined. These definitions are entirely superfluous, in so far as they do not enable us to understand the terms defined, unless we are already familiar with the ideas they are intended to convey. It is probable that Euclid himself did not regard these as real definitions. He probably regarded the notions of "point," "line," "straight line," etc., as primitive notions the meaning of which was clear to every one. The definitions then merely serve to call attention to some of the most important intuitional properties of the notions in question. We will so regard them for the time being We shall have more to say of them presently

Postulates. — Euclid gives us next a set of *postulates*. On account of their historical importance we will give them in full as they appear in the text of HEIBERG [1]

1 *It shall be possible to draw a straight line joining any two points.*

2. *A terminated straight line may be extended without limit in either direction.*

[1] *Euclidis opera omnia*, edited by HEIBERG (Leipzig, 1883–1895). An excellent English edition based on Heiberg's text with critical notes has recently appeared, viz T. L. HEATH, *The Thirteen Books of Euclid's Elements, translated from the Text of Heiberg*, with Introduction and Commentary, 3 vols. (Cambridge, 1908).

3. *It shall be possible to draw a circle with given center and through a given point.*

4. *All right angles are equal.*

5 *If two straight lines in a plane meet another straight line in the plane so that the sum of the interior angles on the same side of the latter straight line is less than two right angles, then the two straight lines will meet on that side of the latter straight line.*[1]

This fifth postulate is the famous so-called *parallel postulate.* On it is made to depend the theorem that through a point not on a given straight line there is only one parallel to the given line.

Axioms. — Euclid now gives a set of *axioms,* "common conceptions of thought," to translate approximately the meaning of the Greek. There are also five of these:

1. *Things equal to the same thing are equal to each other.*
2. *If equals be added to equals, the results are equal.*
3. *If equals be subtracted from equals, the remainders are equal.*
4. *The whole is greater than any one of its parts*
5. *Things that coincide are equal.*

These definitions, axioms, and postulates form the starting point of Euclid's Elements. We may note in passing a very plausible distinction between the axioms and the postulates, which is suggested by this arrangement into sets of

[1] Another discrepancy between the old manuscripts may here be noted. The fifth postulate, as just given, is in some texts given as the eleventh or twelfth axiom.

five. It appears that the axioms are intended to state fundamental notions of logic in general, which may be regarded as valid in any science. The postulates, on the other hand, seem to be intended as primitive propositions concerning space; they are all geometrical.

Criticism of Euclid's Treatment. — We have seen what from a purely logical point of view the starting point of a mathematical science should be. Does this set of axioms and postulates satisfy the requirements? We may at this point dismiss the axioms with the statement that modern criticism is chiefly to the effect that they are too general to be valid in the sense in which the terms involved are now used. As an example, we may call attention to the fact that Axiom 4 (the whole is greater than any of its parts) is not always true in the sense in which the words "whole," "part," and "greater than" are used to-day. We shall return more fully to this on a later occasion.

As to the postulates relating to the fundamental conceptions of space, we must note first that Euclid fails to specify with the necessary precision what terms are to be regarded as undefined. We have already ventured the opinion that he probably regarded such notions as "point," "line," "straight," "length," etc., as primitive notions, the meaning of which is to be regarded as sufficiently clear without any more formal characterization. Is this conception of these notions justifiable? Waiving this question for the moment, we are confronted with the other: Do the postulates satisfy the requirement of a set of unproved propositions; *i.e.* can all the theorems of geometry be de-

rived from them by the methods of formal logic without any further appeal to geometric intuition? We have already stated that Euclid made many tacit assumptions in his derivation of these theorems. He assumes for example without explicit statement that the shortest distance between two points is measured along the straight line joining them. The answer to the last question must then be negative. There remains still another question: What is the logical significance of the postulates? Are they to be regarded as self-evident, necessary truths? This question is at once seen to be closely connected with the first: Are the fundamental notions of "point," "line," "distance," etc., so simple as to have a perfectly clear, precise meaning? We shall devote the next lecture to a discussion which will show that, on the contrary, the connotations of these terms are extremely complex, and that the meaning to be attached to them is by no means clear.

LECTURE II

A NON-EUCLIDEAN WORLD

Logical Difficulties. — The present lecture is to be devoted to an attempt to show that such fundamental notions as straight line, plane, and distance, far from having a precise meaning, are decidedly vague, and subject to a large number of tacit assumptions. We may get some intimation of the difficulties involved, if we try to describe clearly what meaning we do as a matter of fact attach to these terms. Consider, for example, the "distance between two points." How shall we tell what it means? We think first perhaps of measuring the distance, with a foot rule, for example. But that helps us not at all to tell what distance means, since the division of the foot rule into equal intervals (units of distance) already presupposes this notion. And what do we mean by equal distances? This seems easier. The distance, whatever it is, between two points A and B is equal to the distance between two points C and D, if the pair of points A, B can be moved, *without changing their mutual distance*, so as to coincide with the pair C, D. But here we are again in difficulty. The condition "without changing their mutual distance" implies that we know already what

is meant by equal distances. "Well," some one will urge, "let us join the points A, B by a straight line. Then we will say that the distance AB is equal to the distance CD, if the segment of the line AB can by a rigid motion of this segment be made to coincide with the segment joining C and D." Does the notion of "distance between two points" necessarily involve the notion of straight line? What then is the significance of the statement that the shortest distance between two points is measured along the line joining them? And what is a "rigid motion"? It is a motion in which the distance between every pair of points remains unchanged. The vicious circle is apparent.

A New World. — Enough has perhaps been said to show that the connotations of the notion of distance are rather complicated. In order to gain a vivid realization of how serious these difficulties really are, let us construct in our imagination an entirely new world. This world, which it is now proposed to describe, will doubtless appear very fanciful at first; but it will serve a serious purpose. Let us imagine a world inclosed entirely within a large sphere.[1] Suppose that in that sphere the temperature changes from point to point, being a maximum at the center and decreasing gradually until it reaches absolute zero at the boundary. In order to make precise our ideas, let us suppose that if the radius of this sphere is a, and the distance of an object from the center is r, the temperature t is always propor-

[1] This world is described by POINCARÉ, *Science and Hypothesis*, English translation by G. B. HALSTED, p. 49 (The Science Press, New York, 1905).

tional to $a^2 - r^2$; that is, $t = c(a^2 - r^2)$, where c is a constant. Then the value of t is a maximum for $r = 0$, that is, at the center of the sphere; and it vanishes when $r = a$, that is, at the boundary. Moreover, on the surface of any sphere within the given sphere and concentric with it, the temperature is constant. Let us assume also that the inhabitants and material bodies occupying this world are very susceptible to changes in temperature, that they grow larger or smaller in size in direct proportion to the temperature, becoming indefinitely small as they approach the boundary. Suppose also that these changes in size take place instantaneously, so that a body at a given point is always in equilibrium with the temperature at that point.

With this description in mind, we ought to begin to see some of the properties of the geometry which a man living in that world would develop. To him it would seem to be of infinite extent, although to us, viewing it from the outside, it seems to be finite. For if he started to walk toward the boundary, as the temperature fell, his body would grow smaller and his steps gradually shorter, contracting indefinitely as he approached the surface of the bounding sphere. To reach the boundary of his world, he would have to take an infinite number of steps. He would be just as sure that his world was infinite in extent as we are in regard to ours.

The question naturally arises whether he would notice how bodies changed in size when their distance from the center of the world changed. The way in which we usually compare the sizes of objects is to place them side by side, or measure them as with a yardstick If he moved from

one part of his world to another and took objects with him, they would all change at the same time and in the same proportion as he changed. So we may be sure there would be no immediate way of his discovering this law.

"Shortest Lines" are Circles. — In some respects his geometry would resemble our own; but in others it would differ considerably from ours. Suppose, for instance, the man wished to go from his house (H) to his barn (B) by the smallest number of steps (Fig. 1). It is reasonable to suppose that the smallest number of steps would not be taken along the straight line (HB) joining the two places, but rather along a path (say HmB) which swerves toward the center, since his steps would be longer there. In fact, it can be rigorously proven, by the calculus of variations, that the path which gives the smallest number of steps is the arc of a circle which cuts the bounding sphere orthogonally. Such a circle we shall call a "shortest line." Through any pair of points (as H and B) within the sphere there is just one such shortest line. In order to study this world a little more closely, we will take a cross section of it through the center, and try to find the system of circles through a point P, which lie in the plane of this section and cut the bounding circle orthogonally. These circles are very easily

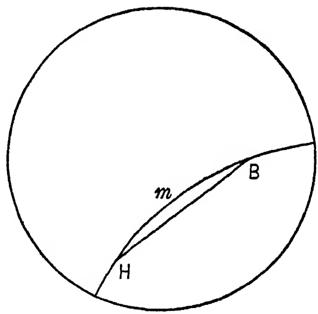

FIG 1

obtained in the following way Through the given point P draw the diameter DD', also a line perpendicular to it through P, which will intersect the bounding circle in a

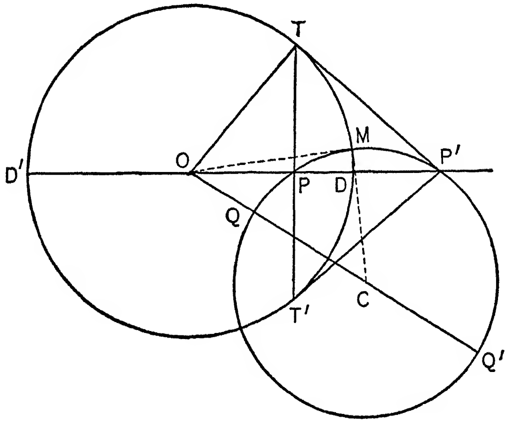

FIG. 2

pair of points, say T and T' (Fig. 2). The tangents to the circle at T and T' will intersect upon the diameter, at a point which we will call P'. *Any circle through P and P' will cut the given sphere orthogonally; and, conversely, any circle cutting the given circle orthogonally and passing through P will also pass through P'.*[1] It follows readily that

[1] To prove this, we notice that OTP' is by construction a right triangle, so that for any point P and the corresponding point P' we always have $OP \cdot OP' = OD^2$. Let QPP' be any circle through P and P , C its center, Q' the other extremity of the diameter through O Then $OM^2 = OQ \cdot OQ' = (OC - QC)(OC + QC) = OC^2 - QC^2 = OC^2 - MC^2$. Whence $OM^2 + MC^2 = OC^2$, and OMC is a right angle. Q. E. D.

through any two points within the sphere there is one and only one shortest line.

A New Assumption. — We will now make another assumption regarding this world, namely that light does not travel along straight lines. We might realize this condition physically by filling the sphere with a gas of which the index of refraction changes from point to point. We will assume that *light travels along the circles which we have just described, i.e. along the shortest lines.* Suppose there is an object at P and a man at Q. According to our last assumption, a ray of light from the object will travel along the arc PQ and reach the eye of the man by the shortest path. If he walked toward the object, keeping it always directly in view, he would move along the arc PQ and arrive at Q by the smallest possible number of steps. We see, then, that *these shortest lines play a rôle in his geometry very similar to that which straight lines play in ours;* in fact, these shortest lines would "look straight" to him. There is one respect, however, in which the geometry of his shortest lines would differ from the geometry of our straight lines.

Euclid's Parallel Postulate. — From his parallel postulate, Euclid derived the theorem, that, through a given point P not on a given line l, there is only one line parallel to l. One way of describing this notion is to draw any line through P cutting l in a point which we will call

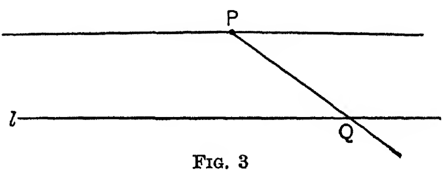

FIG. 3

Q (Fig. 3). If the point Q moves *in either direction* along l, the line PQ will at the same time revolve about the point P. The farther Q proceeds along l, in one of the two possible directions, the nearer will PQ approach a limiting position. Since there are two directions in which Q can move without limit along l, there are two ways of obtaining limiting positions of the line PQ *Euclid's fifth postulate implies that these two limiting positions are the same.*

Parallels in the New World. — Let us see what proposition the people in the "circle world" would have corresponding to the proposition just quoted, if we think of the shortest lines playing for them the rôle of straight lines. We shall see that they would have no use for Euclid's postulate, and that as a statement of a general truth it would even appear ridiculous. Corresponding to the line l, they would have a circle l, meeting the boundary of the world at right angles (Fig. 4). Through a point P, not on l, they would construct another shortest line intersecting l at a point Q, then let Q travel off in either direction along l. For each position of Q there would be a definite circle PQ cutting the boundary orthogonally. Let us see what the limiting positions of PQ would be.

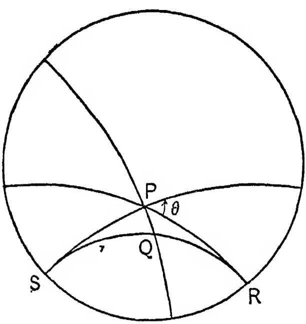

FIG. 4

To the inhabitants of the sphere, Q would appear to be approaching the point R on the boundary, where l inter-

sects it. But since their world would seem infinite to them, and no material object could ever reach its boundary, it would seem to them that the point Q could never reach R. They would consider PR a limiting position of PQ. By allowing Q to travel in the opposite direction, they would obtain a second limiting position PS. These two circles do not, in general, coincide. The angle θ, which they make with each other, is in general so appreciable that the inhabitants would easily observe it. Defining two shortest lines to be parallel, if they are in the same plane and do not meet however far they are produced, we see that the inhabitants of this world would recognize the existence of an infinite number of shortest lines through P and parallel to l, viz. all such lines lying within the angle θ. The two limiting parallels PR and PS we may call the *principal parallels*. On account of these contradictions to euclidean geometry, we might call theirs a *non-euclidean world*.

To avoid a complicated discussion, we have confined our attention to the geometry of a plane section through the center of the sphere. The question might be raised, how do we know that planes in our geometry correspond to planes in this non-euclidean world? It can be shown that the things which correspond to our planes are surfaces of spheres which cut the surface of the bounding sphere orthogonally. But no sphere through the center would cut the boundary orthogonally, unless it degenerated into a plane, which may be thought of as the limiting form of the surface of a sphere as the radius is indefinitely increased.

The Earth placed in this World. — We had a very serious

purpose in describing this world. We will now think of it as being of immense size, but still finite in extent. Imagine our earth placed relatively near the center, so we could examine only a small portion of the world in the immediate neighborhood of the center. The shortest lines through the center are straight lines; the shortest lines through a point very close to the center are approximately straight lines. In fact, we may make the curvature of the shortest lines through a point just as small as we please by placing the point relatively near enough to the center.

An Abstract Science *vs.* a Concrete Application. — Let us pause for a moment to call attention to a certain distinction to the neglect of which is due a large part of the misunderstandings and acrimonious discussions that are so frequent concerning the subject of non-euclidean geometry. The distinction is that between an abstract science and its concrete representation or application. For example, if we conceive of geometry as describing the properties of the space in which we live, that is, the space accessible to our senses, and which we are in the habit of explaining in terms of the sensations of sight and touch, we must remember that we are applying that science to things to which strictly it does not apply. When we talk about points, we are thinking of abstractions, things without size; lines are without breadth, planes are without thickness. We know that such things are not observable in the world accessible to our senses; we can observe only things which approximate to them. Even when by means of carefully made instruments we are able to draw thousands of straight lines

to the inch, we still have things which possess breadth, not the lines of abstract geometry. Think of two lines which make an angle of $\frac{1}{1000}''$. There is no instrument in the world accurate enough to distinguish the angle included between them. Yet we do conceive it abstractly. When we wish to determine whether our abstract science can be applied to our world, the best we can do is to test it, to see if it corresponds to the observed properties of concrete space to *within the possible error of observation.* In other words, there are certain small distances and angles which we conceive of abstractly, but which we cannot realize at all in the concrete application.

The Angle θ. — Now let us return to our non-euclidean world. We have seen that the shortest lines which correspond in the geometry of this world to the straight lines of euclidean space, differ from straight lines by as little as we please in a relatively sufficiently small region near the center of the bounding sphere. Let us consider again the parallels through a given point P to a given shortest line l. We have seen that through P there are two principal parallels, which make at P a certain angle θ. This angle θ decreases as the point P and the line l approach the center of the sphere, and may be made as small as we please by taking P and l sufficiently close to the center. We must remember that "sufficiently close" must be understood relatively to the radius of the bounding sphere. It is perhaps better to put it in this way: *If P and l are at a given distance from the center, the angle θ decreases indefinitely as the radius of the bounding sphere is increased; and*

may be made as small as we please by taking this radius sufficiently large.

Non-euclidean Geometry is Applicable to our Space — Now suppose our earth were at the center of such a sphere, and suppose the radius of the latter is assumed to be so large that in all the space about the earth which is accessible to our observation the angle θ above mentioned is smaller than any instrument of ours can detect. *The observations on the space in which we would thus find ourselves would differ in no particular from those to which we are accustomed; everything would look and feel just the same.* But— and this is the point that is to be emphasized — *the abstract non-euclidean geometry we have above described may be applied to these observations quite as legitimately as the euclidean.* If we keep in mind the distinction between an abstract science and its concrete application, we see there is absolutely no way of telling whether we live in a euclidean or in a non-euclidean world. We do know that the angle θ, if it exists, is so small that we cannot detect it. So, for all practical purposes, euclidean geometry is the most convenient to use, and there can be no doubt that it is the form of geometry which will always be used in practical applications. What we wish to emphasize is that the idea of there being more than one parallel to a line through a given point is in no way inconceivable, and in no way contradicts anything that we observe in our world.

Which is the Correct Geometry? — The question is often asked, and has been very largely discussed in the past hundred years, is the euclidean geometry true, or is the

non-euclidean the true geometry? We see now that such a question has no meaning.[1] It is very much as if we were to ask if the metric system is true; whether it is more correct to measure things in centimeters than in feet and inches. We might very well ask whether it is more *convenient* to measure things in feet and inches than it is to measure them in centimeters. Non-euclidean geometry is much more complex than euclidean geometry, and we shall always find it more convenient to employ the latter in elementary concrete applications of mathematics.

Our present object is attained, if we have succeeded in showing that *our intuitional knowledge of space is not in itself sufficient to characterize with precision and completeness the meaning to be attached and the properties adhering to the fundamental abstract conceptions of geometry*, and, in particular that *Euclid's parallel postulate, however evident centuries of tradition have made it seem to us, is not the only conceivable one to describe the situation in question.* "How, then," it will be asked, "is it possible to build up a valid geometry?"

[1] Cf. Poincaré, *loc. cit.*, p. 39.

LECTURE III

ON THE HISTORY OF THE PARALLEL POSTULATE

Euclid's Attitude toward his Fifth Postulate. — Before attempting to answer the question proposed at the close of the last lecture, it seems desirable to consider briefly the history of the parallel postulate. Euclid's fifth postulate states that *if two lines in a plane are cut by a transversal in such a way as to make the sum of the interior angles on the same side of the transversal less than two right angles, then the two straight lines meet on that side of the transversal.* There is internal evidence to show that, for some reason, Euclid himself did not regard this postulate as being quite so fundamental, or quite so self-evident (if we may use that expression) as his other postulates. For, although he states it with the other postulates, he avoids using it until Theorem 29 of Book I. This theorem says that, *if the given lines are parallel, the sum of the interior angles is two right angles.* He divides his discussion of the exterior angles of a triangle into two parts, as he probably would not have done had he not wished to avoid using the fifth postulate. He first shows, in Theorem 16, that *in any triangle, an*

exterior angle is greater than either of the non-adjacent interior angles. By means of Theorem 16 he proves Theorem 28, that, *if two lines are cut by a transversal so as to make the sum of the interior angles equal to two right angles, the lines are parallel.* For this theorem he did not need to use the fifth postulate. It is the converse of this, namely, Theorem 29, mentioned above, in which he found difficulty and was forced to use the postulate. In Theorem 32 he continues his discussion of the interior angles of a triangle, and shows that their sum is precisely two right angles. This division of the treatment of interior and exterior angles into two parts can hardly be regarded as accidental. It seems to show that Euclid was not entirely satisfied with his fifth postulate, and avoided its use as long as possible.

Ancient Times and the Middle Ages. — The ancient philosophers give us ample evidence that, for a long time after Euclid, it was the fashion among them to discuss the fifth postulate, to try to prove it a consequence of the others, or to replace it by a more satisfactory one; but no essential improvement upon Euclid's treatment resulted for nearly two thousand years. The Dark Ages were as dark in mathematical research as in other things. Not much was done until the time of the Renaissance, about the latter part of the fifteenth century. Then much activity was awakened. This, however, was confined mainly to algebra, such as the solution of cubic and biquadratic equations. Not much was done with geometry, and there is scarcely anything to recall concerning the fifth postulate, until near the end of the seventeenth century.

Wallis, Saccheri, Legendre. — The movement to replace the fifth postulate by a more satisfactory one again attracted attention when taken up, at that time, by JOHN WALLIS (1616–1703), who gave a proof, as he called it, of Euclid's postulate. His proof depends upon the assumption, that *if there is given a triangle in a plane, we can construct a triangle similar to it and of as large an area as we please.* This assumption, however, is practically equivalent to Euclid's. It would not be difficult to show that it is not satisfied in the circle world described in the last lecture. He might indeed have simplified his assumption a little. It would have been sufficient, if it were possible to construct one triangle similar to a given triangle.

GIROLAMO SACCHERI, an Italian Jesuit priest, who wrote about 1733, took a long step in advance. He constructed a new proof, as he supposed, of the fifth postulate, by the method of *reductio ad absurdum*. This form of proof was so well known and so common even in Euclid's time, that it seems strange it had not been employed for this purpose before. He made *an assumption contrary to Euclid's postulate*, and tried to show that it led to an absurdity. He considered two parallel straight lines AC and BD crossed at right angles by a third line AB (Fig. 5). On the lines AC and BD he laid off equal distances AC and BD on the same side of AB,

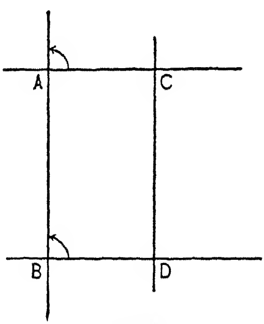

FIG. 5

and joined the extremities of the segments so formed by a fourth line CD. He could show, without the use of the fifth postulate, that the two new angles thus formed were equal. Hence these angles were (1) both right angles, or (2) both obtuse, or (3) both acute. He proved readily that they could not both be obtuse. He then made the assumption that both were acute. He derived property after property — for instance, that the sum of the angles of a triangle is less than two right angles — without coming to any contradiction. He found that, if there were a single triangle in space the sum of whose angles is two right angles, then Euclid's postulate would follow as a consequence from it. The deductions soon became very complicated, however. He supposed that they involved some contradiction, and that Euclid's postulate was therefore established. But he was mistaken. For we know that all the other postulates of Euclid are satisfied by non-euclidean geometry, that is, by the other assumptions made by Saccheri, and the fifth postulate could in no wise follow from them, since it would have to contradict itself. We shall return to this point later. *Saccheri was the first to develop a body of theorems of non-euclidean geometry,* although apparently he did not know that he could not prove Euclid's fifth postulate.

We ought also to mention the work of the French mathematician LEGENDRE (1752–1833). The first edition of his *Elements of Geometry* appeared in 1794. He attempted to rewrite Euclid's Elements in a form as clear and logical as possible. To accomplish this, he made use of some new

postulates; but it was not till 1823, in the twelfth edition of that most popular work, that Legendre claimed he had proved the fifth postulate of Euclid. He was right in thinking that he had proved it, but in proving it he made another assumption, which is rather interesting, since, on account of our habits of thought, it seems so evident and fundamental. He assumed that, *given two half lines issuing from a point O, and any point P in the same plane with them, a straight line can always be drawn through the point P and intersecting both the given half lines.* Why is this postulate not so self-evident as it seems? Let us recall the circle world described in the last lecture. We will take for simplicity the half lines OP, OQ, issuing from the center of the sphere and piercing the boundary at P and Q (Fig. 6). Let R be the given point, and let us see if a shortest line can always be drawn through it, which will intersect both half lines. If we draw the circle through P and Q which cuts the boundary orthogonally, we can easily see that if R lies between the arc PaQ and the boundary PbQ, as at S, no shortest line through R cutting one of the given half lines will ever cut the other half line.

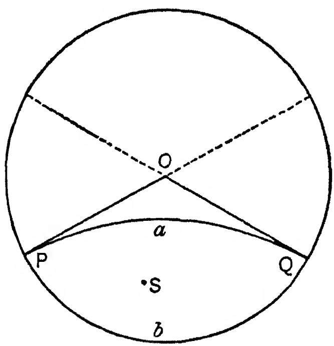

FIG. 6

Gauss. — Among mathematicians recently interested in

non-euclidean geometry was Gauss (1777–1855). In some letters to friends, written near the beginning of the nineteenth century, he said many things that intimate that he was himself for some years in possession of the principles of non-euclidean geometry, but it is doubtful just how extensive his investigations were. In one place he expressed himself to the effect that he was becoming more and more convinced that the fifth postulate of Euclid cannot be proved without an additional assumption. He said he did not wish to publish his results until they were complete, especially since they were of so startling a nature.

Lobatchewsky and Bolyai. — Non-euclidean geometry, as such, appeared first in 1835, when the Russian mathematician, Nicholaus Lobatchewsky (1793–1856), published what he called his Imaginary Geometry. Several articles on the foundations of geometry, which had been published previously, as far back as 1829, gave an indication of the author's genius. Johann Bolyai (1802–1860) also published articles on non-euclidean geometry, in 1832. The two men probably worked independently. They carried the subject far beyond the point which Saccheri had reached one hundred years before. Saccheri, we have seen, had stood on the very threshold of the science, and had even entered the edifice; but either he did not have the courage of his convictions or he did not realize what possibilities lay ahead of him. Lobatchewsky and Bolyai began in the same way. When discussing our ideas of parallelism, in the last lecture, we took a line l and a point P without it, and considered the limiting positions of the line joining the

points P and Q, when Q moved along l without limit in either direction. Euclid assumed that the two limiting positions were the same. Lobatchewsky and Bolyai assumed that they made an angle with each other, which, of course, amounted to saying that there was an infinite number of lines passing through the point P and parallel to the line l.

The Fifth Postulate not Provable. — As we recall the differences between that and our own euclidean geometry, let us not think that it contradicts our experience of actual space in the slightest degree. It merely contradicts our habits of thought. We have seen in the last lecture how this is possible. Another point which we may emphasize at this time is the fact that the set of assumptions which they made are thoroughly self-consistent; starting with those assumptions, we could not expect in any way ever to arrive at a contradiction, and it is therefore evident that *Euclid's postulate cannot be derived from the other postulates.* The problem is now on a par with the squaring of the circle and the trisection of an angle by means of ruler and compass. So far as the mathematical public is concerned, the famous problem of the parallel postulate is settled for all time.

The Consistency of Non-euclidean Geometry. — The question of the consistency or non-consistency of a body of propositions, like those of non-euclidean geometry, is a large one, and we shall have occasion to consider it again in greater detail. For the present, however, it may be of interest to indicate in a rough way how the self-consistency of non-euclidean geometry has been established. For this

purpose, we will construct a sort of dictionary by means of which we can translate propositions of non-euclidean over into euclidean geometry, or vice versa. We may write the equivalent words opposite each other in the following way:

NON-EUCLIDEAN	EUCLIDEAN
Point.	Point.
Straight line.	Circumference of circle perpendicular to a bounding sphere.
Plane.	Surface of sphere perpendicular to a bounding sphere
Angle.	Angle.

The way in which the translation takes place may be shown by the following example. *Theorem in non euclidean geometry:* The sum of the angles of a triangle is less than two right angles. *Corresponding theorem in euclidean geometry* The sum of the angles of a triangle of circular arcs perpendicular to the bounding sphere is less than two right angles. Let us imagine the propositions and theorems of the two geometries all written down, each statement in the one opposite the corresponding statement in the other. Then if we assume that our ordinary euclidean geometry is a self-consistent science, the non-euclidean geometry must also be self-consistent For, corresponding to any contradiction in non-euclidean geometry on the one side, there would have to be a contradiction in euclidean geometry opposite it.

Riemann. — Another kind of non-euclidean geometry was developed by RIEMANN (1826–1866) about 1850. In a very important paper on the assumptions which lie at the basis of geometry,[1] he suggested that there is still a third possibility for a parallel postulate, and hence for a new geometry. Returning again to our figure regarding the limiting position of PQ, Riemann assumed that when the point Q had traveled for some distance along the line l in a given direction, it returned from the opposite direction along the other side of the line, so that the line PQ did not approach any limiting position at all. This assumption contradicts not only the fifth postulate, but also the postulate that a line can be extended indefinitely. It involves the assumption that a line is finite in length. In Riemann's geometry, the sum of the angles of a triangle is always greater than two right angles. Euclidean geometry, then, lies between the other two geometries, forming a limiting case, as it were, for both of them.

Résumé. — It is not our purpose to study the various geometries in detail. We shall be satisfied with what has been said thus far, if it is now possible for us to regard as established that the mere knowledge of space which we get from our senses of sight and touch is not sufficient to characterize completely the abstract space of geometry. We have seen that the starting point of a strictly logical science must be a set of undefined terms and a set of unproved propositions. We have also seen that the simplest

[1] *Ueber die Hypothesen welche der Geometrie zu Grunde liegen*, Habilitationsschrift (Göttingen, 1854).

notions, such as point and line, those which we would prefer to use for our starting point and leave undefined, are by no means so clear as we sometimes think; and therefore, as an immediate consequence, our unproved propositions and those derived from them are by no means as certain as we sometimes think they are. I repeat the question: How, then, is it possible to build up a geometry that is logically valid?

points P and Q, when Q moved along l without limit in either direction. Euclid assumed that the two limiting positions were the same. Lobatchewsky and Bolyai assumed that they made an angle with each other, which, of course, amounted to saying that there was an infinite number of lines passing through the point P and parallel to the line l.

The Fifth Postulate not Provable. — As we recall the differences between that and our own euclidean geometry, let us not think that it contradicts our experience of actual space in the slightest degree. It merely contradicts our habits of thought. We have seen in the last lecture how this is possible. Another point which we may emphasize at this time is the fact that the set of assumptions which they made are thoroughly self-consistent; starting with those assumptions, we could not expect in any way ever to arrive at a contradiction, and it is therefore evident that *Euclid's postulate cannot be derived from the other postulates.* The problem is now on a par with the squaring of the circle and the trisection of an angle by means of ruler and compass. So far as the mathematical public is concerned, the famous problem of the parallel postulate is settled for all time.

The Consistency of Non-euclidean Geometry. — The question of the consistency or non-consistency of a body of propositions, like those of non-euclidean geometry, is a large one, and we shall have occasion to consider it again in greater detail. For the present, however, it may be of interest to indicate in a rough way how the self-consistency of non-euclidean geometry has been established. For this

purpose, we will construct a sort of dictionary by means of which we can translate propositions of non-euclidean over into euclidean geometry, or vice versa. We may write the equivalent words opposite each other in the following way:

NON-EUCLIDEAN	EUCLIDEAN
Point.	Point.
Straight line.	Circumference of circle perpendicular to a bounding sphere
Plane.	Surface of sphere perpendicular to a bounding sphere
Angle.	Angle

The way in which the translation takes place may be shown by the following example. *Theorem in non euclidean geometry*. The sum of the angles of a triangle is less than two right angles. *Corresponding theorem in euclidean geometry:* The sum of the angles of a triangle of circular arcs perpendicular to the bounding sphere is less than two right angles. Let us imagine the propositions and theorems of the two geometries all written down, each statement in the one opposite the corresponding statement in the other. Then if we assume that our ordinary euclidean geometry is a self-consistent science, the non-euclidean geometry must also be self-consistent. For, corresponding to any contradiction in non-euclidean geometry on the one side, there would have to be a contradiction in euclidean geometry opposite it.

Riemann. — Another kind of non-euclidean geometry was developed by RIEMANN (1826–1866) about 1850. In a very important paper on the assumptions which lie at the basis of geometry,[1] he suggested that there is still a third possibility for a parallel postulate, and hence for a new geometry. Returning again to our figure regarding the limiting position of PQ, Riemann assumed that when the point Q had traveled for some distance along the line l in a given direction, it returned from the opposite direction along the other side of the line, so that the line PQ did not approach any limiting position at all. This assumption contradicts not only the fifth postulate, but also the postulate that a line can be extended indefinitely. It involves the assumption that a line is finite in length. In Riemann's geometry, the sum of the angles of a triangle is always greater than two right angles. Euclidean geometry, then, lies between the other two geometries, forming a limiting case, as it were, for both of them.

Résumé — It is not our purpose to study the various geometries in detail. We shall be satisfied with what has been said thus far, if it is now possible for us to regard as established that the mere knowledge of space which we get from our senses of sight and touch is not sufficient to characterize completely the abstract space of geometry. We have seen that the starting point of a strictly logical science must be a set of undefined terms and a set of unproved propositions. We have also seen that the simplest

[1] *Ueber die Hypothesen welche der Geometrie zu Grunde liegen*, Habilitationsschrift (Gottingen, 1854).

notions, such as point and line, those which we would prefer to use for our starting point and leave undefined, are by no means so clear as we sometimes think; and therefore, as an immediate consequence, our unproved propositions and those derived from them are by no means as certain as we sometimes think they are. I repeat the question: How, then, is it possible to build up a geometry that is logically valid?

LECTURE IV

LOGICAL SIGNIFICANCE OF DEFINITIONS, AXIOMS, AND POSTULATES

ONE of the problems which was suggested in the first lecture was: What is the logical significance or the logical character of the definitions, axioms, and postulates in geometry? We are now in a position to give a partial answer to that question. We can at least reject some of the answers that have been advanced by philosophers in the past.

Kant and Mill. — The great German philosopher, IMMANUEL KANT (1724–1804), claimed that the axioms and postulates were *a priori* synthetic judgments imposed upon the mind, without which no consistent or accurate reasoning would be possible. At that time, non-euclidean geometry was already in existence and known by mathematicians to be consistent, but it had not spread rapidly enough for philosophers in general to learn about it. So for a long time they followed Kant, and taught that our conceptions of space were *a priori* judgments, while the mathematicians knew very well that they were not. Only in recent years

have mathematicians and philosophers come near enough together to see that the philosophers were wrong.

Kant, as has been said, claimed that our conceptions of space were *a priori* judgments That this is certainly not so, we have seen. JOHN STUART MILL (1806–1873) claimed that they were experimental facts We have seen also that this is not so. It is true that our experience of space gives us the starting point for our postulates by suggesting what they should be, but our experience of space does not enable us to determine which of several sets of postulates to choose.

Mathematics and Philosophy. — It is interesting to note in passing that, just as in ancient times, mathematics started in conjunction with philosophy, so at the beginning of the renewal of great productivity of mathematics in recent times, namely in the seventeenth and early part of the eighteenth century, the mathematicians were nearly all philosophers, among whom such as PASCAL and LEIBNITZ would be counted. At the beginning of activity, the two sciences (if we may call philosophy a science) were intimately connected; but they soon drifted very far apart, so that at the beginning of the last century the two fields seemed entirely separate, and mathematicians and philosophers were very little interested in each other. In the last thirty years, however, they have been coming very close together again. Finally, it has come to pass that the first real progress in logic and the first essential addition to that subject since the time of ARISTOTLE has been made through the investigations of mathematicians.[1]

[1] RUSSELL, *International Monthly*, vol. 4 (1901), p. 84.

Postulates are Mere Assumptions. — Granting that our geometry need satisfy only our experience of space, and that we may therefore choose a set of non-euclidean or a set of euclidean postulates, as we please, the question still remains as to what those postulates or axioms intrinsically are. So far as we have seen, *they are apparently mere assumptions or agreements concerning the properties of space, which we make at the outset in order to get started.* Suppose we have before us a geometry, consisting of a mass of propositions which for some reason or other we believe to be consistent, and which we wish to rearrange and exhibit in the form of a sequence of logical deductions. We must first choose those terms which we wish to leave undefined and those propositions which we wish to leave unproved. As a practical expedient we would, of course, choose the simpler propositions and try to get the more complex from them. Theoretically, there is no reason why we should not choose the more complex. There is absolutely no restriction upon our choice except that which has already been mentioned: The terms which we leave undefined must be such that every other term we use may be defined in terms of them, and the set of propositions which we leave unproved must be such that we can derive all the others from them by formal logic, without any further appeal to intuition.

An Example. — In order to illustrate these facts, let us choose a few simple propositions from euclidean *plane* geometry, and see what we can derive from them formally.

1. If A and B are points, there exists a line containing A and B.

2. If A and B are distinct points, there exists not more than one line containing A and B.

3. Through a point not on a given line there is one line which does not meet the given line.

4. Through a point not on a given line there is not more than one line which does not meet the given line.

5. Every line contains at least three points.

6. Not all the points are on the same line.

Finally, in order to give us a starting point, we will assume:

7. There exists at least one line.

Let us see what terms are involved in these propositions. We will write them in two columns.

point	is
line	exists
(line) contains (point)	not more than one
(line) through (point)	
(line) meeting (line)	
(point) on (line)	

The words in the right column stand for primitive notions of logic, behind which we shall not try to go. As for the words "one" and "three," we will assume for the present that the meaning of these and of all the positive integers is known. Of the terms in the left column, we are certainly obliged to leave "point" and "line" undefined.

The Notion of a "Class."—The last four terms which remain mean very much the same thing. We may replace them all by a single primitive notion of logic. Let us regard a line as being a *class* of points. By this we mean an assemblage,

collection, totality, or set, of points. If we wish to describe further what we mean by class, we may say that a class of objects is defined whenever a rule or condition is given whereby we can tell of any given object or thing whether it belongs or does not belong to that class. The chairs in this room, for instance, form a class, for each member or *element* of it is uniquely defined. The positive integers form a class; so do the points on a line; so also the presidents of the United States. The class of numbers consisting of all square numbers which end in seven is called a *null* or *empty* class, since there exist no elements belonging to it.

We have left "point" undefined. The points form a certain class which we shall call S. We shall denote the elements of that class by A, B, C, . . . We will think of a line as an *undefined subclass* of these points. By *subclass* we mean a class such that every element in it is an element of the original class S. We will call one of these subclasses an *m-class*. If now we say that two lines meet, we mean that they have a point in common. If we say a point is on a line, it means the same as if we said the point is *of* the line. The ideas of a *class* and *belonging to a class* are notions so fundamental that without them it would be very difficult to proceed at all.

A Miniature Mathematical Science.— Now let us rewrite the propositions stated above in such a way as to exhibit the fact that we are starting with only two undefined terms. We shall thus rid ourselves of any preconceived connotations and any irrelevant ideas, and may proceed to develop our science without attaching any meaning to our words

other than that which is contained in our explicitly stated assumptions. In order to enter into the spirit of this discussion, we must free our minds of any geometrical ideas which may be lurking there. Imagine, if you please, that the things about which we are talking are men, and that the m-classes are committees of men.

1. *If A and B are elements of* S, *there exists an m-class containing A and B.*

2. *If A and B are distinct elements of* S, *there exists not more than one m-class containing A and B.*

3. *Given an m-class a and an element P of* S *not in a, there exists one m-class containing P and not meeting (that is, not having an element in common with) a.*

4. *Given an m-class a and an element P of* S *not in a, there exists not more than one m-class which does not meet a.*

5. *Every m-class contains at least three elements of* S.

6. *Not all the elements of* S *belong to the same m-class.*

7. *There exists at least one m-class.*

Let us see if we can draw any conclusions from these assumptions. From (7) and (5) we immediately deduce that there are at least three elements in S. This is a new proposition not contained among the original assumptions. Let us call these elements 1, 2, 3.

By (1) there exists an m-class containing 1 and 2. By (2) there is not more than one m-class containing these elements. Let us denote this unique m-class by 1 2. By (6) there exists at least one element not contained in 1 2. Call it 4. Then, by (1) and (2) again, we have two new m-classes, 2 4 and 1 4. We had 3 belonging to the class

1 2, and by (5) we have two new elements, say 5 and 6, belonging respectively to 1 4 and 2 4. Let us therefore change our notation for the three m-classes to 1 2 3, 1 4 5, 2 4 6. The elements 3, 5, 6 are all distinct, for if any two of them, say 5 and 6, were not distinct, the two m-classes to which they belong, 1 4 5, 2 4 6, would have the pair of elements 4 and $5=6$ in common, so that by (2) they would be the same m-class, which is contrary to the hypothesis. By (1) there is consequently a new m-class 1 6, which by (5) contains an element 7 distinct from all the others.

We will now make use of assumption (3). Take for a the m-class 2 4 6, and for P the element 1. By (3) there is an m-class containing 1 which does not meet 2 4 6, and therefore cannot contain any of the elements 3, 5, 7. For if it did, it would meet 2 4 6. By (5) it contains besides 1 at least two other elements. Let these be 8 and 9. We will stop here with this *theorem: The class* S *contains at least nine distinct elements.*

Already, we have some idea of how we may proceed *to reason formally with a set of absolutely undefined elements, concerning which we have made a certain number of fundamental assumptions,* the results being devoid of all content except such as is contained in the set of assumptions.

LECTURE V

CONSISTENCY, INDEPENDENCE, AND CATEGORICALNESS OF A SET OF ASSUMPTIONS

Consistency. — The question now arises, are the fundamental assumptions which we made during the last lecture logically *self-consistent*, or could we arrive at contradictory conclusions? This is, of course, a question of the greatest importance. The only test for the consistency of a body of propositions is that which connects with the abstract theory a *concrete representation* of it. We are dealing here with a collection of symbols. If we can give them a concrete interpretation which satisfies, or appears to satisfy, all our assumptions, then every conclusion that we derive formally from those assumptions will have to be a true statement concerning this concrete interpretation. As an illustration, consider the set of assumptions which we have been discussing. They were taken originally from geometry. Suppose instead of regarding the elements of S as mere symbols devoid of meaning, we interpret them as points in a plane, and m-classes as straight lines in a plane. Then every one of our assumptions is satisfied by euclidean geometry.

If now we grant that euclidean geometry is consistent, every proposition which we derive by means of formal logic will be true of every m-class, and hence of every concrete representation of them. For, if they were not consistent, there must be some contradiction in euclidean geometry. The fundamental question is thus answered only by reference to a concrete representation of the abstract ideas involved, and it is such concrete representations that we wished especially to avoid. At the present time, however, no absolute test for consistency is known. A simpler consistency proof for this set of assumptions will appear presently.

A New Assumption. — We will now add another assumption to the set of seven with which we began to build up our miniature science in the last lecture. Let us purposely choose one which is not satisfied by euclidean geometry.

8. *No m-class contains more than three elements of* S.

We can no longer use euclidean geometry as a concrete representation of our science; for each of the triples which we found, namely, 1 2 3, 2 4 6, 1 6 7, 1 4 5, 1 8 9 constitutes a whole m-class, and, under this new assumption, a line could not contain more than three points. If we wish to have a concrete representation, we may, as suggested before, imagine the m-classes to be committees of men. Each committee must, then, consist of three men. Moreover, there cannot be more than four m-classes containing the element 1, namely, 1 2 3, 1 4 5, 1 6 7, and 1 8 9. For, suppose there were a fifth, say 1 AB. The elements A and B must be distinct from each of those which we have been consider-

ing, otherwise, by assumption (2), $1AB$ could not be a new m-class. Since A and B must be distinct from each of the old elements, $1AB$ could not meet 2 4 6. But, by (3) and (4), there is one and only one m-class containing 1 which does not meet 2 4 6, and we already have 1 8 9, which satisfies those conditions. Therefore, a fifth m-class containing the element 1 cannot exist. Let us find all the m-classes containing 2. We already have 1 2 3 and 2 4 6. We might take 2, 5, and 7 together to form a new m-class 2 5 7. But, in that case, we would be unable to find an m-class containing 2 which did not meet 1 4 5. The existence of an m-class 2 8 9 is contradicted by the fact that we already have 1 8 9. No two men, for example, can serve together on more than one committee, on account of assumption (2). Instead of taking 2 5 7, however, let us take 2 5 8. This will allow us to take also 2 7 9. Proceeding in the same way, we might take as additional m-classes containing 3, 3 4 8, 3 5 7, 3 6 9, but a contradiction would arise from this choice. If, however, we take for the new m-classes containing 3, 3 4 9, 3 5 7, 3 6 8, we shall be able to select another new m-class containing 4, namely, 4 7 8, also one containing 5, namely 5 6 9. Clearly, if we choose our m-classes in the manner indicated, we cannot obtain more than these twelve without violating our assumptions. Let us arrange them in a table in the order in which we obtained them.

1 2 3	2 4 6	3 4 9	4 7 8	5 6 9
1 4 5	2 5 8	3 5 7		
1 6 7	2 7 9	3 6 8		
1 8 9				

This arrangement of the nine digits into sets of three constitutes a proof that the set of eight assumptions with which we have been working is consistent. If by the class S we mean the digits from one to nine inclusive, or any set of nine distinct things designated by these nine digits, and if by an *m*-class we mean the above triples, then every one of the assumptions is satisfied. Every consequence which we derive from these assumptions must be true of these twelve triples. Corresponding to any contradiction in the propositions derived from our fundamental assumptions, we would find a contradiction concerning these triples, which is impossible. Moreover, if we take away one of our fundamental assumptions, the remaining seven are certainly consistent.

Before leaving this example, let us notice that the above set of triples is not the only set which will satisfy our eight assumptions. When we chose the *m*-class 2 5 8, we might have taken 2 5 9 instead. In that case, we would have constructed a table of twelve triples as follows:

123	246	348	479	568
145	259	357		
167	278	369		
189				

We shall have occasion to return to this table very soon.

The set of assumptions upon which any mathematical science is built up should ordinarily possess three fundamental properties. The first and most important property is consistency. It is always required. No mathematical science is possible without a consistent set of fundamental assumptions.

Independence.—The second fundamental property sought for in a set of assumptions is *independence*. By this we mean that none of the assumptions can be derived as a formal logical consequence from the others. For example, it can be shown that none of the eight assumptions which we have just been discussing follows as a logical consequence of the other seven. That set of assumptions is therefore independent as well as consistent. If we wish to make a clear distinction between a theorem (*i.e.* a proposition derived from assumptions) and an assumption, we must be sure that our unproved propositions are completely independent. If, however, we are not greatly concerned with this distinction, then the question of independence is not of so much importance. In this sense, then, independence is not an essential requirement, but merely, from some points of view, a desirability.

We saw that the first seven assumptions had a complete concrete representation in ordinary euclidean geometry. On the other hand, the eighth assumption contradicts ordinary euclidean geometry, since it implies that there are not more than three points on a line. If it were possible to derive the eighth as a consequence of the first seven, it would have to express a property of ordinary euclidean geometry. *The eighth assumption is therefore independent of the others.*

In general, an independence proof is constructed in the following way: Let there be given a set of assumptions of any nature, numbered 1, 2, 3, $\cdots$ n, and let it be required to prove that Assumption No. k is independent of all the others. We must find one concrete representation for

which all the assumptions, except No. k, are satisfied and for which No. k is not true. The exhibition of such a concrete representation constitutes an independence proof of the Assumption k.

Categoricalness.—A third fundamental property of a set of assumptions often, though not always, desirable, is *categoricalness.* To explain what we mean by this, let us consider our miniature science again We did not have the slightest knowledge of what the elements of S were, for they were left completely undefined. The question arises whether, with such an abstract formulation of a problem, merely by making assumptions of that sort concerning undefined elements, we can completely characterize the class to which those elements belong. Suppose, for example, we make a set of assumptions for euclidean geometry, in which we regard point and line as unknown things, and treat them as empty symbols devoid of all content, except such as is implied by the assumptions. Can we, in this abstract way, completely characterize what we mean by euclidean geometry, that is, can we choose our assumptions so that the propositions which follow from these will be abstractly equivalent to the propositions of euclidean geometry alone, and not also of some essentially different science? As a matter of fact, this can be done.

A while ago, we saw that there were two sets of triples of nine digits, which satisfy all the assumptions which lie at the basis of our miniature science. It is easy to see why there must be at least two such sets. When we found the elements 8 and 9, we obtained them both together.

We might have called the first one 9 and the second one 8. The only difference between the two sets is simply that 8 and 9 are interchanged. The two sets are not essentially different; they differ only in notation We will say that two classes $S = (A, B, C, \cdots)$ and $s = (a, b, c, \cdots)$ are not essentially different, if we can establish a correspondence between the elements of the two classes such that to the element A in S corresponds the element a in s, to B corresponds b, and so on, a correspondence with the property that every statement which is true of a system of elements chosen from S and is derived from a given set of assumptions, will be equally true of the corresponding elements in s. If we have any statement concerning any system of elements of S, we may obtain a corresponding statement of the corresponding system of elements of s, simply by replacing the large letters in the statement by the corresponding small letters. Any two systems of elements which can be put into *reciprocal one-to-one correspondence with preservation of properties as to a set of assumptions,* in the way just described, are said to be *isomorphic* with respect to this set of assumptions. When a set of assumptions satisfies the condition that *any* two concrete applications of it are simply isomorphic, the set of assumptions is said to be *categorical.*

A Categorical *vs.* a Non-Categorical Set. — Here a question arises. Is it always desirable that the set of assumptions from which we build up a mathematical science be categorical? The answer depends upon the object we have in view. It is an advantage to keep the set non-categorical

as long as possible, for the reason that, if we build up a science on a set of assumptions which is non-categorical, there will be more than one system of things which satisfies the assumptions, that is, there will be at least two essentially distinct concrete representations of it. There will thus be a gain in generality. Consider, for example, the eight assumptions of our miniature science. The first seven are non-categorical, for they are satisfied by the set of nine numbers, also by the points on a plane, and the points on a plane cannot be put into reciprocal one-to-one correspondence with the nine numbers. In general, *by starting with a non-categorical set of assumptions, we can develop a part of several abstract sciences at the same time.* We obtain, in that way, a theory which may be part of a great many different theories. The great gain in generality thus obtained is so obvious that examples are hardly necessary. The principle is illustrated by our miniature science. If we work with all eight assumptions, instead of the first seven only, we obtain a set of propositions which are not all true of euclidean plane geometry, and which apply to a much narrower field of science, namely the arrangements of nine things into triples in a certain way. All theorems derived from the first seven only, however, are equally true of the system of triples and of ordinary euclidean geometry.

A Set of Assumptions for any Branch of Mathematics. — This miniature mathematical science of triples illustrates what is done on a larger scale in setting up a set of assumptions for any branch of mathematics. First, it is necessary

to choose the terms which are to be left undefined in such a way that every other term may be defined in terms of them. Secondly, it is necessary to choose the propositions which are to be left unproved in such a way that all the other propositions of the science can be derived by the methods of formal logic from these chosen ones. The resulting science is thus purely abstract, and may have several concrete representations. If the set of assumptions is categorical, any two of these representations will be isomorphic.

Advantages of Abstract Formulation. — Such an abstract formulation is essential in order to exhibit with precision the logical connections of the science. What we have just said shows, however, that such a formulation has also a *great unifying power*, in that it exhibits clearly the analogies existing between several sciences. This fact makes the abstract treatment of great importance, quite aside from its logical significance.

Pasch and Peano. — The abstract formulation of mathematics seems to date back to the German mathematician, MORITZ PASCH (1843–). At any rate, he was the first to study in detail the axioms concerning the order of points on a straight line and to state clearly the assumptions involved in the idea of "betweenness." We shall have occasion to return to this later. But to the Italian GIUSEPPE PEANO (1858–) belongs the credit of developing this point of view systematically. His idea, which he began to elaborate about 1880, is to put the whole of mathematics on a purely formal basis, and for this purpose he invented a symbolism of his own. In 1893 he began the publication of a "For

mulario di matematica," which is a synopsis of the most important propositions of the different branches of mathematical science, with their demonstrations, expressed entirely in terms of symbolic logic. He is the founder of a school of thinkers in Italy, whose ideas and writings are just beginning to receive the attention which they deserve.

The Democracy of Mathematics.—An immense change in the point of view toward the foundations has been brought about since this abstract formulation was put forward. Some one, I think it was VAILATI, has suggested that this change is very similar to that which a nation undergoes when it changes from a monarchic or aristocratic form of government to a democracy. The point of view fifty years ago was very largely that the foundations of mathematics were axioms; and by axioms were meant self-evident truths, that is, ideas imposed upon our minds *a priori*, with which we must necessarily begin any rational development of the subject. So the axioms dominated over mathematical science, as it were, by the divine right of the alleged inconceivability of the opposite. And now, what is the new point of view? The self-evident truth is entirely banished. There is no such thing. What has taken the place of it? Simply a set of assumptions concerning the science which is to be developed, in the choice of which we have considerable freedom. The choice of a set of assumptions is very much like the election of men to office. There is no logical reason why we should not choose the more complex propositions; but as a matter of fact we usually choose the simpler, because it is easier to work with them. Not all propositions reach the

high position of assumptions; they are elected for their fitness to serve, and their fitness is very largely determined by their simplicity, by the ease with which the other propositions may be derived from them.

The Assumptions in the Rôle of Definitions. — We have seen that it is possible by means of a set of assumptions to characterize any mathematical science completely, as far as the logical relations are concerned; that this is true whenever a set of assumptions is categorical. As soon as we realize this, we see that, although we start with terms which are entirely undefined, when we have imposed a sufficient number of assumptions to make the whole set categorical, the assumptions will, in a sense, act as definitions of the undefined elements. As far as their logical character is concerned, *the unproved propositions play the rôle of disguised definitions.* With regard to this thoroughly abstract point of view, we can see what truth there is in that rather humorous definition given by RUSSELL: "Mathematics is the science in which we never know what we are talking about, nor whether what we say is true." There is a sense in which Russell's statement is strictly true. We do not know what we are talking about, because the things about which we are talking are entirely undefined; and we do not know whether what we say is true, because we cannot know whether the assumptions regarding these undefined terms are true. Every mathematical theorem is, logically, of the form, *if P* is true, then *Q* is true. Pure mathematics, in the abstract form, says nothing concerning the truth of its propositions. To give Russell's serious definition of mathematics, he said

that *mathematics is the class of all propositions of the form, P implies Q.*[1] It may be criticized as too broad. It includes more than is usually included by the term "mathematics."

Pure and Applied Mathematics. — If we adopt the point of view of PEANO and RUSSELL, all *pure* mathematics is abstract. Any concrete representation of such an abstract science is then a branch of *applied* mathematics. Geometry, for example, as a branch of pure mathematics, consists, then, simply of the formal logical implications of a set of assumptions. Whenever we think of geometry as describing properties of the external world in which we live, we are thinking of a branch of applied mathematics in the same sense in which analytic mechanics is a branch of applied mathematics. We need not quibble over this distinction. The important thing is to recognize that *there exists an abstract science underlying any branch of mathematics,* and that *the study of this abstract science is essential to a clear understanding of the logical foundations.*

A Message for Teachers. — It was stated at the beginning of these lectures that the discussion would have an important message for teachers of elementary mathematics. We have been concerned so far mainly with non-euclidean geometry and the elucidation of the purely formal and abstract point of view. I hope it will not be thought that I advocate the introduction of non-euclidean geometry or of abstract methods into our high school courses. Nothing, in my opinion, could be more absurd. But the facts concerning the logical significance of undefined terms and

[1] RUSSELL, *The Principles of Mathematics*, p. 1.

unproved propositions have a very important bearing on elementary teaching. The great majority of our textbooks in geometry begin with a set of formal definitions and a few axioms and postulates, then follows immediately the sequence of formal propositions. To attempt formal definitions of such things as point, straight line, plane, etc., is scientifically unjustifiable and pedagogically undesirable. One of the things we wish to impart to our pupils is a clear understanding of the force of a definition, to teach them to learn the meaning of an unknown or vaguely understood word by defining it with precision in terms of words of which the meaning is known. Is it going to help a high school pupil to gain a clear notion of the nature of a definition by giving him at the very outset of his study the following as a sample?

"A straight line is a line of unlimited extent such that any part of it will coincide with any other part, if the extremities of the two parts are made to coincide."

This is very much as though we were to say to our boys: "Here is an example of a definition: A *boy* is a male of the genus homo who has not reached the age of an adult. Is it not clear how this statement tells you the meaning of 'boy'? And is it not interesting?" Just as absurd it is to expect the pupil to become interested in and to understand the nature of a demonstration by giving him as first examples alleged proofs of propositions which appear to him so obvious that he can see no reason for a proof. Such a procedure merely confuses him and stifles his interest.

Two General Pedagogical Principles. — Two general principles may be stated:

No FORMAL *definition of any term should be given that can not be defined in terms of ideas obviously simpler than the term to be defined.*

No FORMAL *proof of any proposition should be attempted which seems obvious to the pupil without proof.*

It may seem to some a waste of time to dwell so long on the obvious. And yet the great majority of our textbooks appear to make it necessary to call attention to the pedagogical absurdities referred to above. It is difficult to see why so many of our textbook writers still incorporate them in their books. It may be that they have a feeling that their work would be scientifically unsound, if they did not formally define and prove everything. But we have seen, on the contrary, that such an attitude is quite as absurd scientifically as it is pedagogically. It is a logical necessity that some terms remain undefined and that some propositions remain unproved. "But," they may say, "is it not desirable to reduce the number of undefined terms and the number of unproved propositions to a minimum?" No! Pedagogically it is very undesirable, and scientifically it is not necessary. To a mature mind the problem of reducing to a minimum the number of undefined terms and of rendering the set of unproved propositions independent is interesting and important; to the mind of the high school pupil the problem has no meaning. By all means let the number of formally undefined terms and the number of what we may call prelim-

inary propositions (*i.e.* propositions formally unproved) be large. Let us remember that our primary object is not to teach our pupils to *know* geometry, but rather to lead them to *think* geometry. This can be done only by arousing their interest in geometric figures and problems and leading them to think about them *in their own way*, at first. The pupil's own thoughts must be gradually led into the formal mode of reasoning; he is not likely ever to learn to think geometrically by being required to repeat the thoughts of another in a form that must in the nature of the case appear to him at first as artificial and unnatural.

LECTURE VI

CLASS. CORRESPONDENCE. NUMBER

Resume. — The last lecture closed what may be regarded as the preliminary or introductory part of this course. The object of these introductory lectures has been to make clear the nature of the problems to be discussed and the point of view from which it is proposed to approach them. The description of an imaginary world served to show that the meaning popularly attributed to certain fundamental concepts (such as distance and straight line) lacks precision, that the axioms and postulates of geometry cannot be regarded as self-evident truths, and that indeed the ordinary euclidean geometry is by no means the only science that will serve to describe the properties of space, — in other words, that our intuitive knowledge of space is not sufficient to determine completely the fundamental propositions of geometry.

It thus became apparent that *a purely logical treatment of geometry implies a purely abstract treatment.* Beginning with the observation that it is impossible to give formal definitions of every term, or to give formal proofs of every proposition without becoming involved in a vicious circle,

it was seen that the starting point of any mathematical science must be a set of undefined terms and a set of unproved propositions (assumptions) concerning them. The science then consists of the formal logical implications of the latter. These considerations were illustrated by the discussion of a miniature mathematical science, by means of which it was possible also to exhibit the properties of *consistency*, *independence*, and *categoricalness* of a set of assumptions. With these general ideas in mind, we are now ready to begin a more systematic discussion of various fundamental concepts of mathematics, and will begin with the notion of *class*, which will lead directly to that of *cardinal number*.

The Notion of Class. — In the abstract formulation of any mathematical science the notions of *class* (or *set*), and of *belonging to a class* are fundamental. Another very fundamental notion is that of *correspondence* between the elements of two classes, which has already been used in the discussion of the categoricalness of a set of assumptions. These notions are primitive concepts of logic, into the meaning of which we do not inquire further. We may first consider these notions without particular reference to mathematics. Is there anything which we can discover or discuss about a class, if we do not know anything concerning the individual elements of the class? Is there any way we can compare two classes without making any use of the nature of the individual elements of which they are composed? It seems that there is just one thing we may do. We may possibly be able to tell whether the number of elements in each class is the same. This leads to a very

important question: What is meant by the "number of elements"? Throughout our discussion we have hitherto supposed that we are familiar with the natural numbers, that is, the ordinary positive integers. If we have two classes in which the number of elements is finite, we may count them to see if the number is the same. But suppose we wish to compare two classes, one of which consists of all the points on a line a foot long, the other of all the points on a line three inches long. We cannot count the number of elements in each. Is there, then, no means of comparison?

One-to-one Correspondence. Cardinal Number. — There is, in fact, a simple means of comparison which does not involve the operation of counting; and which does not presuppose the notion of number at all. It depends on the notion of correspondence. Consider, for example, a regiment of soldiers. Ordinarily each soldier of the regiment has one gun, and every gun belongs to one soldier. The number of guns and the number of soldiers is then the same; we know that without doing any counting. This method stated precisely is as follows: Let there be given two classes. If we can establish a law or rule whereby with every element of one class is paired a certain element of the other class, and vice versa, we say that the two classes can be placed *into a one-to-one reciprocal correspondence*, or are *equivalent*. In this case we say also that the two classes have the same *cardinal number*.[1] This involves an

[1] This relation between two classes is also expressed sometimes by saying that they have the same *power*, a translation of the German *Machtigkeit* introduced by G. CANTOR.

extension in the use of the word "number." It preserves the same sense in all cases where the original meaning of the word "number" applies, that is, when the number of elements is finite, but it offers a means of comparing classes containing an infinite number of elements.

An Example. — Let us consider, for example, any two segments, AB and $A'B'$, of straight lines (Fig. 7) which we suppose to be in the same plane (but not on the same line). *The class of all points on the segment AB may be put into one-to-one reciprocal correspondence with the class of all points on the segment $A'B'$.* To prove this, let O be the point of intersection of the lines AA' and BB'. If, then, to any point P of AB we make correspond the point P' of $A'B'$ which is on the line OP, we have established a correspondence of the desired kind between the class of points on the segment AB and the class of points on the segment $A'B'$.[1] *The cardinal numbers of these two classes are therefore equal*, by definition.

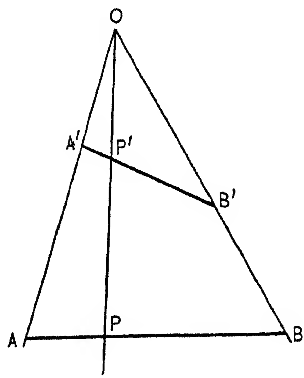

FIG 7

It follows almost immediately that *the cardinal number of the class of points on any line segment AB is the same*

[1] If AA', BB' are parallel, we replace the line OP by the line through P parallel to AA'. The argument then applies also to this case.

as the cardinal number of the class of points on any line segment CD contained in AB. To prove this we need only assume an auxiliary line segment $A'B'$ in a plane with AB, but on a different line The points of AB may then be placed into one-to-one correspondence with the points of $A'B'$ as already described, and these in turn placed into correspondence with the points of CD (cf. Fig. 8). But the points of CD form a part only of the points of AB. *We have thus an example of a class which has the same cardinal number as a part of itself.*

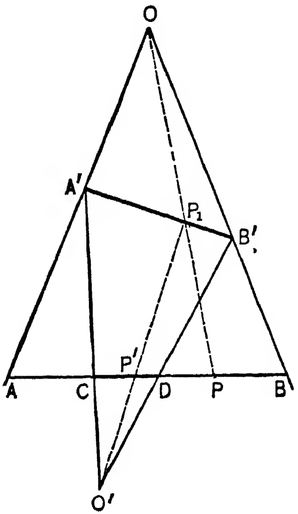

FIG. 8

Another Example. — Another example of this kind of class is furnished by the class of all positive integers. It is readily seen how the latter class may be put into one-to-one correspondence with the class of all even positive integers, by writing the classes as follows: —

$$\begin{array}{cccccccccccc} 1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & 10, & 11, \cdots. \\ | & | & | & | & | & | & | & | & | & | & | \\ 2, & 4, & 6, & 8, & 10, & 12, & 14, & 16, & 18, & 20, & 22, \cdots. \end{array}$$

To any integer a of the first class corresponds the integer $2a$ of the second; and, conversely, to any integer $2b$ in the second corresponds the integer b in the first. *The class*

of all positive integers therefore has the same cardinal number as the class of all even positive integers, and the latter is only a part of the former.

Infinite Classes. — We have thus seen two examples of classes — viz., the class of all points on a line segment, and the class of all positive integers — which have the property that there exists a part of the class which can be put into one-to-one correspondence with the whole class. This property is characteristic of the classes called infinite. In fact, we may now formulate the following definition: —

A class C *is said to be infinite if it contains a part* C′ *which can be put into one-to-one correspondence with* C. *A class which is not infinite is said to be finite.*

It should be noted that a *part* C′ of a class C is any subclass of C such that there exists at least one element of C which is not also an element of C′. From this definition follows readily the following important property of infinite classes: —

If a class C *is infinite, the class* C_1 *obtained by removing from* C *any element is also infinite.*

For by hypothesis C contains a part C′ which may be put into one-to-one correspondence with C. Let a be the element of C which is removed to obtain the class C_1, and let a' be the element of C′ which corresponds to a in the correspondence existing between C and C′. The class C_1' obtained by removing from C′ the element a' is then a part of C_1 which is in one-to-one correspondence with C_1. The class C_1 is, therefore, infinite by definition.

It follows from this theorem that *an infinite class can-*

not be exhausted by removing its elements one at a time. This connects our definition of an infinite class with the more popular conception of this term.

We return now to the question: *What is the number of elements in a class? What is a cardinal number?*

Older Attempts to define Cardinal Number. — Philosophers have attempted to define a cardinal number by the process of enumeration. Such an attempt is, however, doomed to failure at the outset, since it presupposes a knowledge of the thing to be defined. For, to enumerate the objects in a class is merely to put these objects into one-to-one correspondence with the numbers 1, 2, 3, . . . n. The impossibility is still more obvious when the attempt is made to introduce psychological elements. "Counting," we are told, "consists of successive acts of attention; the result of such a succession is a number." In other words, the number six is the result of six acts of attention. The vicious circle is obvious.

But even if some such method were possible, it has at least two disadvantages. In the first place, it could apply only to finite classes, whereas we have to deal so often with infinite classes, that it is desirable to have the word "number" apply also to them. In the second place, it presupposes that the elements of the class in question can be arranged in order (since counting involves such an arrangement), and there may be classes of which this is impossible. In any case, the notion of "order" is a new concept which is not essentially involved in the notion of number, and which it is therefore desirable to avoid in defining a number.

Abstract Definition of Cardinal Number. — The notion of one-to-one correspondence offers a means of giving a precise formulation of the notion of cardinal number which satisfies all logical requirements and which is entirely free from the disadvantages just mentioned Given a class C, the set N of all classes which may be put into one-to-one correspondence with C (that is, which are equivalent to C) is a set with the property that any two classes which compose it have, by definition, the same cardinal number. *The cardinal number of the class* C — *or the number of elements in* C — *is then simply a symbol* N *representing the set of all classes that can be put into one-to-one correspondence with* C. All classes are thus associated into sets N. Any two classes in the same set have the same cardinal number, any two classes of different sets have different cardinal numbers. The class of all cardinal numbers then is a class of symbols, each of which represents one and only one of these sets N. These symbols serve merely to recognize two classes as belonging or not belonging to the same set N.

This definition of cardinal number makes no distinction between finite and infinite classes; but applies to any class whatever. Moreover, it does not make use of the notion of order. It is only subsequent to this definition that we may define such a relation between cardinal numbers. Before doing this, however, we will consider the notion of "order" itself. In a later lecture we shall then see how to define the notions of "greater than" and "less than" applied to cardinal numbers. There also will be found the answer to the question which has probably already suggested itself:

F

Is there more than one infinite cardinal or are not all infinite classes equivalent?

The Whole Greater than any of its Parts. — It was stated in the first lecture that modern criticism of Euclid's axioms is mainly to the effect that they are too general; and attention was called to the axiom: The whole is greater than any of its parts. We see now why this axiom is too general. We have seen that the number of elements in an infinite class is equal to the number in one of its parts. The axiom in question is characteristic of finite wholes.

LECTURE VII

ORDER. DISCRETE SEQUENCES

Relations and Operations. — We have hitherto considered the notion of a class simply as such, without assuming it to have any special properties. We have seen how it is possible with no further assumptions to define the notion of a cardinal number and to distinguish between finite and infinite classes. The classes with which mathematics is concerned, however, are of many kinds — classes of points, of numbers, etc., with certain characteristic properties. It is our purpose to consider in detail some of these classes and to see how they may be characterized abstractly by their properties. These properties are in general of two kinds. Either the elements of the class in question are subject to certain *relations*, or they are subject to certain *operations*. As examples of the first kind of properties we have the relation of *order* (of two numbers one is greater than the other, of two points one is to the left of the other, etc.), the relation of *collinearity* (three points are, or are not, on the same line), etc. As examples of the second kind we mention the operations of addition, multiplication, etc., applied to numbers.

Linear Order.—We will begin with the discussion of the relation of *order*, more precisely of so-called *linear order.* This will give us our first example of a familiar notion characterized abstractly by a set of assumptions. We will represent this relation, which we assume as *undefined*, by the symbol $<$. This symbol may be read "precedes," "less than," "above," "older than," etc.; but care must be had in using these words not to attribute to the symbol any of their possible connotations which are not implied by the assumptions presently to be made regarding this symbol.

The expression $a=b$ (read. a "equals" or "is the same as" b) indicates that the elements a, b in question may be interchanged in the discussion. The expression $a\neq b$ (read: a "is distinct" or "is different from" b) indicates that a and b may not be interchanged. The relation $<$ is then characterized by the following *fundamental assumptions:*[1]

Given a class C and a relation $<$; let a, b, c be any elements of C.

O_1. *If $a\neq b$, then either $a<b$ or $b<a$.*

O_2. *If $a<b$, then $a\neq b$.*

O_3. *If $a<b$ and $b<c$, then $a<c$.*

It follows as a theorem that *if a, b are any two elements of* C, *we have either $a=b$, or $a<b$, or $b<a$.*

Consistency.—Let us test this set of assumptions for consistency, independence, and categoricalness. As for consistency, we need only give one concrete interpretation

[1] This set of assumptions is due to E. V. Huntington; cf. reference at bottom of p. 71. The notion of order had previously been studied by Vailati, Padoa, Pasch, and others.

of the class C and the symbol $<$ for which these assumptions are satisfied. There are a large number of examples of that sort, of which we need mention only the natural numbers, with $<$ meaning "less than," or the points of a line, with $<$ meaning "precedes."

Independence. — As regards independence, we can prove that each of the three postulates is independent of the other two. In order to show that the first assumption is independent of the last two, that is, that it is not derivable as a consequence of them, we must find a concrete representation of the last two for which the first is false. Every consequence of the last two will have to be a valid statement of the particular concrete system which we employ, and if the first contradicts that, then this assumption cannot be a logical consequence of the last two. To give such an example, consider the class H of all human beings in history. Interpret $<$ to mean "is an ancestor of." Assumption O_2 reads now, if a is an ancestor of b, a is not b. O_3 says, if a is an ancestor of b, and b is an ancestor of c, then a is an ancestor of c. Both statements are, of course, true. But it is not true, as O_1 now states, that, if a is not b, then either a is an ancestor of b, or b is an ancestor of a. To prove that the second postulate is independent of the other two, let C be the class of natural numbers and interpret $<$ to mean "equal to or less than." Then the first and third postulates are satisfied by the natural numbers, whereas the second is not. An independence proof for the third postulate is obtained at once by merely interpreting the symbol $<$ to mean "different from." If we wish another example for

this purpose, which is not quite so briefly disposed of, we may take five points, A, B, C, D, E, equidistantly distributed upon the circumference of a circle. Let us interpret the relation $<$ to mean that the arc joining in a counter-clockwise direction the two elements in question is less than a semi-circle. The first and second postulates are satisfied by this system of points. The third is not satisfied. We have now proven that *the given assumptions defining linear order are mutually independent.*

Categoricalness. — Is this set of assumptions categorical? That is, can any two classes which satisfy it be made isomorphic with each other, so that, if of any two elements a, b in the first class a precedes b, then in the second class the element corresponding to a will precede the element corresponding to b? It is easy to see that *the set is not categorical.* For the set of assumptions is satisfied by any finite set of integers arranged in their natural order, as well as by the infinite set of all integers. Moreover, the natural numbers and the points on a line both satisfy the set, as we have already remarked, and it is not difficult to show that these two classes cannot be put into a one-to-one correspondence, although they are both infinite.

Types of Order. — This not only shows that our set of assumptions for linear order is non-categorical, but it leads us to the investigation of various *types of order*, a subject of prime importance in discussing the foundations of mathematics. A class C satisfying the conditions for linear order will be denoted by (C, $<$), and is called an *ordered* class. If two ordered classes (C, $<$) and (C′, $<$) are such that they

can be put into one-to-one correspondence in such a way that, if of any two elements of C, a first precedes a second, then of the two corresponding elements in C′, the first will also precede the second, the two classes are said to be *ordinally similar* or to belong to the *same type of order*.

We shall be able to give only a brief outline of some of the important types of order. We shall see that the ordinary real number system of algebra, the rational numbers by themselves, the integers by themselves are all examples of different types of order, and that they can be completely distinguished, so far as abstract properties are concerned, as types of order One great achievement of mathematical research of recent years has been the proof that the class of all real numbers and the class of all points on a line may be completely characterized simply by the notion of order, without the notion of magnitude or measurement.

Definitions. — In discussing types of order, we will begin with some definitions[1]

Given two elements a and b of an ordered class (C, $<$); an element x of C such that $a < x$ and $x < b$, is said to be *between* a and b.

If we have $a < x$, and no element of C is between a and x, then x is called the *immediate successor* of a.

If we have $x < a$, and no element of C lies between x and a, x is said to be the *immediate predecessor* of a.

[1] We are following here, in a general way, the article by E. V. Huntington, "The Continuum as a Type of Order," *Annals of Math.*, Vol. 18, 1904–1905, p. 151.

If there exists an element x in C which precedes all other elements in C, x is called the *first* element of C.

If there is an element in C which is preceded by all other elements of C, that element is called the *last* element of C.

With respect to the existence of first or last elements in an ordered class (C, $<$), we can divide all classes (C, $<$) into four kinds. The class C may be such that (1) it contains neither a first nor a last element, (2) it may contain a first but no last, (3) it may contain a last but no first, or finally, (4) it may contain both a first and a last element.

Some Examples. — It seems desirable to precede any further discussion by some examples which will illustrate the great variety of order types with which we are confronted.

Consider the sequence of positive integers from 1 to 10 in their natural order:

(1) $$1, 2, 3, 4, 5, 6, 7, 8, 9, 10.$$

This class has a first element; it also has a last element. Every element except the last has an immediate successor; every element except the first has an immediate predecessor.

Again, consider the unlimited sequence of positive integers,

(2) $$1, 2, 3, 4, 5, \cdots$$

in their natural order. This ordered class has a first element, but no last. Every element has an immediate successor; every element except the first has an immediate predecessor.

The unlimited sequence of negative integers in their natural order,

(3) $$\cdots, -5, -4, -3, -2, -1,$$

has no first element, but has a last. Every element has an immediate predecessor, every element except the last has an immediate successor.

The unlimited sequence of positive and negative integers and zero in their natural order,

(4) $$\cdots -3, -2, -1, 0, 1, 2, 3, \cdots$$

has no first and no last element; but every element has an immediate successor and an immediate predecessor.

Now let us consider the class of all positive rational numbers in their order of magnitude, that is, let $<$ be interpreted to mean "less than." This class has no last element; it has no first element. Also no element has an immediate predecessor, since between any two rational fractions there exist other such fractions. For the same reason no element has an immediate successor. It is evident that all classes thus far considered belong to different types of order.

The Same Class may represent Different Types of Order. The class last considered, viz. that of the positive rational numbers, will serve to show that *the elements of certain classes may be arranged in a linear order in several essentially different ways.*

We considered this class above with reference to the usual order of magnitude, $<$ meaning simply "less than." We will now consider two other ways of ordering this class.

The class in question consists of all fractions of the form $\frac{m}{n}$, m and n being positive integers having no common factor. Suppose now we arrange this class of fractions as follows:

(5) $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \cdots; \frac{2}{1}, \frac{2}{3}, \frac{2}{5}, \frac{2}{7}, \frac{2}{9}, \cdots; \frac{3}{1}, \frac{3}{2}, \frac{3}{4}, \frac{3}{5}, \frac{3}{7}, \frac{3}{8}, \cdots;$

Here we take first all rational numbers whose numerator is 1 in ascending order of their denominators, then all those whose numerator is 2 in ascending order of their denominators, then all those whose numerator is 3, and so on. In other words, with this meaning of $<$, of two fractions a, b, with different numerators, $a < b$, if the numerator of a is less than the numerator of b; of two fractions a, b, with the same numerator, $a < b$, if the denominator of a is less than the denominator of b. This arrangement clearly satisfies the fundamental assumptions O_1, O_2, O_3. With this ordering the class has a first element, but no last, and every element has an immediate successor. But some elements have immediate predecessors, while others (such as $\frac{1}{1}, \frac{2}{1}, \frac{3}{1}$, etc.) have not. This type of order is different from any of those previously described.

A third method of ordering this class is obtained by writing the fractions in a rectangular array as follows:

$$
(6) \quad \begin{array}{llllll}
\frac{1}{1}, & \frac{1}{2}, & \frac{1}{3}, & \frac{1}{4}, & \frac{1}{5}, & \cdots\cdots \\
\frac{2}{1}, & \frac{2}{3}, & \frac{2}{5}, & \frac{2}{7}, & \cdots\cdots & \\
\frac{3}{1}, & \frac{3}{2}, & \frac{3}{4}, & \cdots\cdots & & \\
\frac{4}{1}, & \frac{4}{3}, & \cdots\cdots & & & \\
\frac{5}{1}, & \cdots\cdots & & & & \\
\cdots\cdots & & & & &
\end{array}
$$

and then arranging them in order by reading them diagonally as indicated. We thus obtain the arrangement:

$$(7) \qquad \frac{1}{1}, \frac{2}{1}, \frac{1}{2}, \frac{3}{1}, \frac{2}{3}, \frac{1}{3}, \frac{4}{1}, \frac{3}{2}, \frac{2}{5}, \frac{1}{4}, \frac{5}{1}, \frac{4}{3}, \ldots$$

The relation $<$ in this case means "precedes" in the order in which the fractions are thus obtained. It is readily seen that this arrangement of the fractions exhibits them as ordinally similar to the sequence of positive integers 1, 2, 3, 4, 5, 6, ..., a result of far-reaching importance. We shall return to this later. We may note in passing, however, that it proves the remarkable theorem: *The cardinal number of all positive rational numbers is equal to the cardinal number of all positive integers.*

Discrete Sequences. — It is our purpose to characterize the types of order mentioned and certain other types of importance in mathematics by further assumptions. To this end we will first give the additional assumptions to characterize the type of order known as a *discrete sequence.* Of this we shall find several subtypes. A discrete sequence is characterized by the following set of assumptions in addition to those we have already made:

D_1. *Dedekind's postulate. If* C_1 *and* C_2 *are any two non-empty subclasses of an ordered class* C, *such that every element of* C *belongs either to* C_1 *or to* C_2, *and such that every element of* C_1 *precedes every element of* C_2, *then there exists an element* X *in* C, *such that every element which precedes* X *belongs to* C_1 *and every element which follows* X *belongs to* C_2. In other words, there is an element X in C which actually brings about the division into two classes. It

may be either the last element in C_1 or the first element in C_2.

D_2. *Every element of* C, *unless it be the last, has an immediate successor.*

D_3. *Every element of* C, *unless it be the first, has an immediate predecessor.*

As an example, consider the class of integers arranged in their usual order of magnitude, negative, zero, positive, thus,

$$\cdots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \cdots.$$

Let us divide it into two classes in any way we please, say at 3. It is quite evident Dedekind's postulate is satisfied. The other two assumptions are also satisfied. If we begin to write the integers with, say, -2, there is a first, but no last element. If we take the original sequence and break it off at, say, 5, there is no first element, but there is a last. The example given shows that this set of six assumptions O_1, O_2, O_3, D_1, D_2, D_3 is *consistent.*

Dedekind's Postulate. — The assumption D_1 is of particular interest, as it describes a property of ordered classes, which is of great importance. Its bearing will become clearer if we consider an example of an ordered class for which it is not satisfied. Consider the class C of all positive and negative integers, and consider them ordered as follows:

$$1, 2, 3, \cdots; \ \cdots -3, -2, -1;$$

that is, the symbol $<$ is to mean that any positive integer precedes any negative integer, whereas of any two positive

integers, or of any two negative integers, the smaller precedes the greater. Assumptions O_1, O_2, O_3 are clearly satisfied with this interpretation of $<$; clearly also D_2 and D_3 are satisfied; assumption D_1, however, is not. For if we let C_1 be the class of all positive integers, and let C_2 be the class of all negative integers, every element of C belongs either to C_1 or to C_2, and every element of C_1 precedes every element of C_2. But, there exists no element X of C which brings about this division, since C_1 has no last element and C_2 has no first. This example proves, incidentally, that *D_1 is independent of O_1, O_2, O_3, D_2, D_3.*

In this connection it may be noted that the ordered class (5), on page 74, satisfies assumptions O_1, O_2, O_3, D_1, D_2, but does not satisfy D_3. The latter is then independent of the previous ones. The same example, with the sense of $<$ reversed, proves the independence of D_2.

Mathematical Induction — As an example of a theorem derived from this set of assumptions we may mention the following, on which the principle of *mathematical induction* depends:

If a and b are any two elements of a discrete sequence C, *and $a < b$, and s_1 is the immediate successor of a, s_2 the immediate successor of s_1, s_3 of s_2, etc., then among the set $s_1, s_2, s_3, \cdots$ occurs the element b.*

To prove this, suppose the theorem were not true. Let C_1 consist of $a, s_1, s_2, s_3, \cdots$ and all the elements which precede a, and let C_2 consist of the remaining elements of C. C_1 has then no last element; for every s_i precedes b, if b is not among the s_i. Hence by D_1, C_2 must have a first

element, which in turn by D_3 has an immediate predecessor, which would have to be the last element of C_1. This involves a contradiction.

The connection of this theorem with the principle of mathematical induction is obvious. This principle is a simple consequence of the fact that the sequence of positive integers $1, 2, 3, \cdots, n, n+1, \cdots$ forms a discrete sequence.

Types of Discrete Sequences. — There are four types of discrete sequences: progressions, regressions, unlimited sequences, and finite sequences, which are defined as follows:

1. Type ω. A *progression* is any discrete sequence which has a first element and no last.

2. Type $^*\omega$. A *regression* is any discrete sequence which has a last element and no first.

3. Type $\omega + {}^*\omega$. An *unlimited discrete sequence* is one which has neither a first nor a last element.

4. Type of *finite discrete sequences,* that is, those which have both a first and a last element.

It is easy to see why a discrete sequence possessing a first and a last element must be finite. Suppose we take away the elements of the sequence one at a time, beginning with the first. The last element is one of the successors of the first element, by the theorem proved above. By taking away all the successors of the first element, we can finally exhaust the sequence.

LECTURE VIII

THE SEQUENCE OF CARDINAL NUMBERS. DENUMERABLE CLASSES. DENSE CLASSES. CONTINUOUS CLASSES

The Sequence of Cardinal Numbers. — It will be recalled that in Lecture VI we defined the cardinal number of a class C to be a symbol representing the class N of all classes which are equivalent to C. This symbol attaches to each class of the class N, and serves to recognize two of these classes as belonging to N. We have not as yet, however, ordered the class of all cardinal numbers; in particular, we have not defined what is meant by saying that one cardinal number is greater than or less than another.

In the case of finite cardinals this definition might simply be as follows: The cardinal number of a class C is less than the cardinal number of a class C′ if the class C is equivalent to a part of C′. This definition will not, however, apply to the cardinals of infinite classes, since in this case C may be equivalent to a part of C′ and at the same time be equivalent to the whole of C′. In order, therefore, that the definition may apply also to infinite cardinal numbers, we formulate it as follows: —

The cardinal number of a class C *is less than that of a class* C′, *if* C *is equivalent to a part of* C′ *and* C′ *is not equivalent to any part of* C.

It can be proved, moreover, that if a class C is equivalent to a part of C′, and C′ is also equivalent to a part of C, then C and C′ are equivalent, that is, have the same cardinal number.[1] It follows at once that of two cardinal numbers, a, b, either $a = b$, or $a < b$, or $b < a$ ($<$ meaning "less than"). This shows that assumptions O_1 and O_2 are satisfied by this definition of $<$. Moreover, it is evident that O_3 is also satisfied. So that by this definition *the class of all cardinal numbers, finite and infinite, is arranged in linear order.*

We obtain thus first of all the progression of finite cardinal numbers in their natural order:

$$1, 2, 3, 4, 5, 6, \cdots$$

The question as to the existence of more than one infinite cardinal now demands an answer.

Denumerable Classes. — An infinite class is said to be *denumerable,* provided it can be put into one-to-one correspondence with a progression, or in particular with the set of positive integers. It has already been shown that the set of all even integers satisfies this definition. A perhaps more remarkable example of a denumerable class is the set of all rational numbers. In the last lecture, we showed how the correspondence with the progression of natural

[1] This is known as BERNSTEIN's theorem, for proof, see BOREL, *Leçons sur la théorie des fonctions*, Note I, Paris, 1898; or RUSSELL, *Principles of Mathematics*, p. 306.

numbers might be established by arranging the (positive) rational fractions in a rectangular array and reading it diagonally. These classes then have the same cardinal number. It is usually denoted by ω, and is called a "denumerable infinity."

Not all classes are denumerable. A good example of one which is not is the class of all positive non-ending decimals less than one. Let us assume that this class can be put into one-to-one correspondence with the natural numbers. Imagine the whole class written down *in the required order*,

$$\begin{array}{lll} 1. & — & 0\,a_1a_2a_3a_4a_5a_6 \cdots \\ 2. & — & 0.b_1b_2b_3b_4b_5b_6 \cdots \\ 3. & — & 0.c_1c_2c_3c_4c_5c_6 \cdots \\ \cdot\cdot & \cdots & \cdots \quad \cdots \quad \cdots, \end{array}$$

where the letters represent digits It is, then, possible to find a non-ending decimal greater than zero and less than one which is not contained in the above list, and which we will write as follows:

$$0.m_1m_2m_3m_4m_5m_6m_7 \cdots \cdots$$

To form it, we need only take m_1 different from a_1, m_2 different from b_2, m_3 different from c_3, and so on The new decimal will differ from the first decimal in the list at least in its first digit, it will differ from the second at least in the second digit, from the third in at least the third digit; in short, it will differ from the nth decimal in at least the nth digit. Since there are ten different digits, there are nine different ways of selecting each m. This proof was given by CANTOR in 1891.

The Cardinal of the Continuum. — The cardinal number of this class is therefore different from the cardinal number of a denumerable class. It is called the *cardinal number of the continuum*, and is generally denoted by $\mathfrak{c}$. It follows at once from the discussion that we have $\omega < \mathfrak{c}$. It can be shown that ω is the first infinite cardinal; also that there exists an unlimited number of cardinals greater than $\mathfrak{c}$. Are there any cardinals between ω and $\mathfrak{c}$? Mathematicians are still awaiting the answer.

Dense Classes. — In the last lecture we characterized what is meant by saying that an ordered class is discrete. We now turn our attention to a second type of ordered class, which is equally important. An ordered class C is said to be *dense*, if, in addition to assumptions O_1, O_2, O_3, it satisfies the following assumption:

H. If a and b are any two elements of C, *there exists an element of* C *between a and b.*

A simple example of a dense class is the class of all positive rational numbers, arranged in their natural order. If $\frac{m}{n}$ and $\frac{l}{k}$ are any two numbers of this class, it is easy to show that the rational number $\frac{m+l}{n+k}$ is between $\frac{m}{n}$ and $\frac{l}{k}$. Another familiar example is the class of points on a line; between any two such points there exists another point of the line.

Continuous Classes. — Finally we should characterize what is meant by a continuous class. An ordered class is said to be *continuous* if it is *dense* (Assumption *H*) and satisfies

Dedekind's postulate (Assumption D_1). As an example, we may mention the class of all points on a straight line. This class is clearly dense Moreover, it satisfies Dedekind's postulate. For this requires simply that if the points of the line be divided into two (non-empty) subclasses C_1, C_2, such that every point of the line belongs either to C_1 or to C_2, then there exists a point X which effects this division Another example consists of the class of all real numbers. In fact, we are in the habit of representing these numbers by the points of a line.

The class of all rational numbers furnishes an example of a class which is dense, but not continuous. For let us divide the class of all rational numbers into subclasses as follows: Let C_1 consist of all rational numbers $\frac{m}{n}$ for which $\frac{m^2}{n^2} < 2$, and let C_2 consist of all the remaining rational numbers The classes C_1 and C_2 then satisfy the conditions for Dedekind's postulate; but since there exists no rational number whose square equals 2, the class C_2 consists of all rational numbers $\frac{m}{n}$ such that $2 < \frac{m^2}{n^2}$. It is readily seen that C_1 has no last element, and that C_2 has no first element. Assumption D_1 is therefore not satisfied. Any discrete sequence is an example for which assumption D_1 is satisfied, but which is not dense.

Fundamental Segment. Limit.—The distinction between a continuous class and one that is merely dense is so vital to a clear understanding of many problems connected with the foundations of mathematics, that it seems desirable to exhibit it in a slightly different form.

A part S of a dense class C is said to be a *fundamental segment* of C, if it has the following two properties.

1. If a is any element of S, any element which precedes a is an element of S.

2. S has no last element.

As an example, let C be the class of all rational numbers in their natural order, $<$ meaning "less than." Let S consist of all those numbers of C, whose square is less than 2. It is then readily seen that S is a fundamental segment.

Given a dense class C and a fundamental segment S of C, *if there exists* in C an element y, such that every element of C which precedes y is an element of S, y is said to be the *limit* of the fundamental segment S.

In the example considered S has no limit. If, however, we let C be the class of all real numbers and S the part of C consisting of all real numbers whose square is less than 2, then S is again a fundamental segment. But in this case it has the limit $\sqrt{2}$ in C. A continuous class may, therefore, be described as *any dense class every fundamental segment of which has a limit in the class.* It will be noticed that this method of description is equivalent to the preceding one.

The Linear Continuum.—We have given as examples of continuous classes the points on a line and the totality of real numbers, and have observed that these two classes are *ordinally similar.* Are our assumptions sufficient to characterize completely this type of order? They are not, as the following example will show.

Let C be the class of all points in and on the boundary of

a square. Let us define order in this class as follows (Fig. 9): Of two points in the square at different distances from the side OY, the one nearer OY shall precede the other; of two points at the same distance from OY (*i.e.* on a line parallel to OY) the one nearer OX shall precede the other. It is then readily verified that the points of this square satisfy all our assumptions for a continuous class (viz. O_1, O_2, O_3, D_1, H).

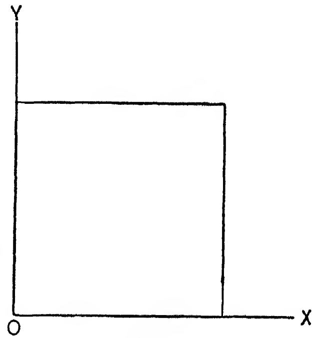

FIG 9

To characterize the type of order to which the points of a line belong, and which is called a *linear continuum*, we will have to add another assumption, or better, replace the assumption of density by another known as the *postulate of linearity*.

If an ordered class C contains a denumerable subclass R such that between any two elements of C there is an element of R, the class C is said to be *linearly continuous*, provided it also satisfies Dedekind's postulate. Take for the class C the set of all real numbers, for the subclass R all the rational numbers This subclass is denumerable, and it is such that an element of it exists between every pair of real numbers, which we may choose to select in the class C. The real numbers therefore constitute a *linear continuum*, according to this definition. It is not difficult to see that the elements of the square above mentioned do not

form a linear continuum. If there existed a subclass R such that between every pair of points in the class of the square there always existed an element of R, there would have to be an element of R for every point on the base line OX of the square. It can be proved that the set of points on a line segment is not denumerable. In fact, if we take the side of the square to be of length unity, and assume that the points of such a side are equivalent to the class of all non-ending decimal fractions between 0 and 1,[1] we have already proved this proposition (p. 81). We found, namely, that the set of non-ending decimals between 0 and 1 is not denumerable.

Cantor's *Mengenlehre*. — We will now leave for the present the subject of ordered classes. The results we have given are practically all due to GEORG CANTOR, who developed them and many others in his so-called *Mengenlehre* (theory of classes), beginning in 1872.[2] We should note particularly that we have succeeded in characterizing abstractly by means of sets of assumptions the following types of order:

1. *Discrete sequences* (Assumptions $O_1, O_2, O_3, D_1, D_2, D_3$), *e.g.* the type of the *integral numbers*.

2. *Denumerable dense classes* (Assumptions O_1, O_2, O_3, H,

[1] It should be noted that any terminating decimal may be written as a non-terminating decimal. Thus $0.24 = 0.23999$

[2] CANTOR's original articles appeared mostly in the *Mathematische Annalen*, *e.g.* Vols. 15, 17, 20, 21, 23, 46, 49. For an exposition in English reference may be made to the articles by E. V. HUNTINGTON, *Annals of Mathematics*, Vols 6, 7; and to the Treatise by W. H. and G. C. YOUNG, *The Theory of Sets of Points*, Cambridge, 1906.

with assumption of being denumerable), *e.g.* the type of the *rational numbers.*

3. *Continuous classes* (Assumptions O_1, O_2, O_3, D_1, H). In particular, *linearly continuous classes* (Assumption H replaced by assumption of linearity), *e.g.* the type of *all real numbers* and the *points on a straight line.*

We are thus making progress in our problem of characterizing the fundamental conceptions of mathematics. We turn now toward the characterization of the number systems of algebra. To do this, we must first discuss the notion of an *operation.*

LECTURE IX

GROUP. NUMBER SYSTEM

Class and Operation. — We have considered the notion of a class by itself as a fundamental notion. We have also discussed the notion of class in connection with the relation of order We will now discuss the notion of class in connection with the idea of *operation.* In order to see clearly just what are the abstract concepts which lie at the basis of a number system, we may first examine an important example of a class in which a single operation has been defined.

Given a class C of which the elements are denoted by a, b, $\cdots$; what are we to understand by an operation upon the elements of the class? We say that an *operation* o upon the elements a and b is defined, if, corresponding to the elements a and b and to a certain order of those elements, there exists a certain third thing c. Here again the notion of correspondence is the central one. The new thing c which is associated with, or corresponds to, the given elements in the given order, is called the *result* of the operation, and we write $aob = c$, or $boa = c'$, according as the order of the elements is a, b, or b, a.[1] If, for example, the

[1] In speaking of order in this connection we have no reference to the notion of order discussed in the previous lectures. We are not

given elements are the numbers 3 and 5, in the order first 3, second 5, and the operation is division, the corresponding result is the number $\frac{3}{5}$. If the elements and the operation had been the same, but the order reversed, the result would have been different, namely $\frac{5}{3}$ If the operation had been addition, the results would have been the same, irrespective of the given order That is, $3+5=8$, and $5+3=8$. Whenever the result aob is equal to the result boa, the operation o is said to be *commutative* (with respect to a and b).

The result c may or may not belong to the class C of the given elements a and b. In the above examples, if we take C to be the class of all positive integers, then the result $3+5$ belongs to the same class as the original elements, but the results $\frac{3}{5}$ and $\frac{5}{3}$ do not belong to that class, and it is necessary to go outside the given class to find the resulting element c.

Definition of a Group. — Closely connected with the idea of operation is the notion of a group with respect to an operation. We will say that a class C is a *group with respect to an operation* o which is supposed to operate between any two elements of C, if the following four assumptions are all satisfied:[2]

supposing the class C to be ordered. We merely wish to call attention to the formal difference between aob and boa.

[2] This definition of a group is substantially that given by L. E. Dickson, "Definitions of a Group and a Field by Independent Postulates," *Transactions of the American Mathematical Society*, Vol. 6 (1905), p. 199.

G_1. *If a and b are in C, then aob is in C.*

G_2. *If $a, b, c, \cdots$ are elements of C, the result of operating upon the elements a and boc, in the order named, is the same as the result of operating upon aob and c, in the order named. That is, $ao(boc) = (aob)oc$.* This is the so-called *associative law.*

G_3. *There exists in C an element i, such that $aoi = ioa = a$, for every element a of C.*

G_4. *There exists in C, corresponding to any element a, another a' such that $aoa' = i$.*

The element i is called the *identity* or the *identical element* of the group. The element a' is called the *inverse* of a.

If we take for the class C the system of ordinary real numbers, or the system of rational numbers, or the system of all integers, and for the operation o the operation of addition, we find that all these assumptions are satisfied. For if $a, b, c, \cdots$ are elements of C, (1) $a + b$ is in C, (2) $a + (b + c) = (a + b) + c$, (3) there exists a number (0) such that $a + 0 = 0 + a = a$, and (4) there exists another a' such that $a + a' = 0$, that is, corresponding to each number, there is another which is the negative of it. *The set of real numbers or the set of rational numbers or the set of all integers, therefore, each forms a group with respect to the operation of addition.* If we take for the class C the real number system, and for the operation that of multiplication, we shall find that the conditions for a group are again satisfied, *except in one particular.* The product of two numbers is always in C, the associative law holds, and there exists a

number corresponding to the identical element i, namely the number 1, since $1 \times a = a \times 1 = a$. For every number a there is another a' such that $aa' = 1$, *except for the number zero.* There is no number which multiplied into 0 gives 1.

It is easy to prove from these assumptions that *in any group there exists only one identical element* i, also that *there is but one inverse for each element.* The assumption that there are two of either of these readily leads to a contradiction.

An operation which is not in general commutative may be so in special cases. Division, for example, is not in general commutative, since if a and b are not equal, $\frac{a}{b}$ is not equal to $\frac{b}{a}$. But if a and b are equal and distinct from 0, the operation of dividing one by the other is commutative, since it always leads to the same result, namely 1. A group in which the operation o is commutative throughout is called a *commutative* group.

A Geometrical Group. — As an example of a group in which occur non-commutative operations, let us consider the rotations of an equilateral triangle ABC about its center or its lines of symmetry so as to transform the triangle into itself. By transforming the triangle into itself, we mean moving it in such a way as to bring the vertices A, B, C back to their original positions, except for a possible interchange or permutation of the letters A, B, C. In this case the elements of the class C are the three rotations of the triangle in its plane about its center through angles of 120°, 240°, and 360°, and the three rotations of the triangle

about its medians through an angle of 180° (Fig. 10). After each of these movements the triangle is superposed upon, or congruent with, its original position. If we take for granted the fact that these six movements are the only possible ones which transform the triangle into itself, we see at once that, if two of the rotations be performed successively, the result is equivalent to a single rotation belonging to the same class. The six rotations therefore satisfy the first postulate, namely that the class of rotations forms a group with respect to the *operation* of combining them. The associative law also holds. The identical element is represented by the rotation about the center through 360°. The inverse of each rotation exists; corresponding to any one of the rotations about the center through an angle of $\alpha°$ there is another rotation about the same point through an angle of $(360 - \alpha)°$ which combined with the original rotation is equivalent to the identical element or the rotation through 360°. Corresponding to each of the three operations of turning the triangle over through 180°, there is a repetition of the turning which brings the triangle into its original position. Each of these turnings is therefore its own inverse. That the operation is not in general commutative may be shown very easily. Let the medians through A, B, C, be lettered a, b, c, respec-

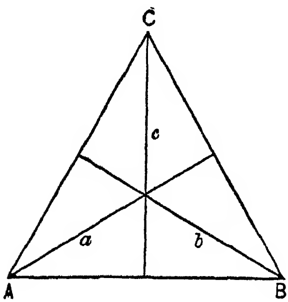

FIG. 10

tively. A rotation through 180° about a puts B into C, C into B, and A into A. The result may be conveniently represented by the symbol $\begin{pmatrix} ABC \\ ACB \end{pmatrix}$ A similar rotation about c now puts A into B, B into A, and C into C, and may be represented by $\begin{pmatrix} ABC \\ BAC \end{pmatrix}$. The result of performing these two rotations in succession is to put A into B, B into C, C into A, or, briefly, it is equivalent to the substitution $\begin{pmatrix} ABC \\ BCA \end{pmatrix}$, which is at once seen to be equivalent to a rotation of 120° about the center. If, however, we change the order of the rotations, performing first the rotation about c, and then the rotation about a, we obtain the result $\begin{pmatrix} ABC \\ CAB \end{pmatrix}$, which is equivalent to a single rotation through 240° about the center. The operation of combining these rotations is therefore not commutative.

Importance of Group Concept.—If we were asked to state the most important single concept which lies at the basis of mathematics, next in importance to the fundamental notions of class and correspondence, we would mention the notion of group The set of all movements of a rigid body in space forms a group. That is, any such movement followed by any other such is equivalent to a single movement of a rigid body. This group is of fundamental importance in elementary geometry. We introduce the notion of group at this point, however, in order to define by means of it what is meant by a number system, in the abstract and most general sense. Later, we will see to what extent the concept of a general number system must be restricted in

order to obtain the system of numbers used in ordinary algebra.

Definition of a Number System. — A *number system* consists of a class N in which two undefined operations, which we will denote by $+$ and $\times$, exist and operate subject to the following three assumptions:

N_1. *N is a group with respect to the operation* $+$. We will denote the identical element with respect to $+$ by i_+, or 0.

N_2. *N is a group with respect to the operation* $\times$, *except that no inverse of* 0 *is required.* The identical element with respect to $\times$ is denoted by $i_\times$, or 1. The third postulate connects the two operations, and is usually called the *distributive law.*

N_3. *If* a, b, c, *are any elements of* N, *we have*

$$a \times (b + c) = a \times b + a \times c, \text{ and } (b + c) \times a = b \times a + c \times a.$$

Observe that in this definition nothing has been said concerning the number of elements which compose the number system N. The number of elements may be finite or infinite. There exist number systems of both kinds. We have also said nothing about the operations $+$ and $\times$ which would imply that they are commutative or not commutative. There are classes which satisfy the postulates, and hence form number systems in this abstract general sense, for which the operations $+$ and $\times$ are not commutative. A number system in which the operations $+$ and $\times$ are both commutative is called a *commutative number system* or a *field.*

The class of all rational numbers forms a group with respect to addition ($+$); it also forms a group with respect

to multiplication ($\times$), except that the inverse of 0 does not exist. The distributive law (N_3) also holds. *The rational numbers, therefore, form a number system* in the sense just defined. Moreover, it is clear that this number system is *commutative.*

It should be noted that the *inverse operations* of *subtraction* ($-$) and *division* ($\div$) may be defined in terms of $+$ and $\times$ respectively. Indeed, the element $a-b$ is defined as the element x such that $b+x=a$; the element $a\div b$ ($b\neq 0$), as the element y such that $b\times y=a$. And such elements x and y always exist in a class N satisfying N_1 and N_2.[1]

A Finite Number System. — Examples of non-commutative number-systems will be given later. We may at this point, however, give a simple example of a number system consisting of only a finite number of elements. Let the elements of the class N in question be the five digits

$$0,\ 1,\ 2,\ 3,\ 4.$$

Let the "sum" ($+$) of any two of these elements be the ordinary sum of the two numbers, if this sum is less than 5; and, if this ordinary sum is equal to or greater than 5, let it be the smallest remainder (positive or zero) obtained by dividing the ordinary sum by 5. Thus:

$$1+2=3,\ 0+3=3,$$
$$1+4=2+3=0;\ 2+4=1, \text{ etc.}$$

[1] What we have defined is more precisely *right-handed* subtraction and division. *Left-handed* subtraction and division is defined similarly by reversing the order of b and x, and b and y. In a commutative number system this distinction is unnecessary.

Further, let the "product" ($\times$) of any two of these elements be defined as the ordinary product, if the latter is less than 5; and, if this ordinary product is equal to or greater than 5, let it be replaced, as before, by the smallest remainder after division by 5. For example:

$$1 \times 3 = 3 \quad 2 \times 2 = 4.$$
$$2 \times 3 = 1. \quad 4 \times 4 = 1. \quad 3 \times 3 = 4.$$

With these definitions the assumptions for a number system are all satisfied. As to the existence of an inverse, we may, for example, say that with respect to addition, the inverse of 1 is 4, since $1 + 4 = 0$. With respect to multiplication, the inverse of 4 is 4, since $4 \times 4 = 1$.

Such a number system is called *modular*, the *modulus* in this case being 5. A modular number system may be defined similarly for any modulus which is a prime number.

Implications of the Definition.—It will be of interest to inquire briefly just how much is implied by the assumptions for a number system. We note first that the only elements, the existence of which is directly assumed, are the identical elements 0 and 1 with respect to addition and multiplication respectively. As far as these assumptions are concerned, moreover, no further elements are necessary. For if we assume $1 + 1 = 0$, $0 \times 1 = 1 \times 0 = 0$, we shall find all the assumptions satisfied. This number system is, indeed, simply the modular system with modulus 2. Let us leave out of consideration this case, however. Since the system forms a group with respect to addition, the elements $1 + 1$, $(1 + 1) + 1$, $[(1 + 1) + 1] + 1$, $\cdots$ are all elements of

the system. Let us *assume* that all elements obtained in this way, by the successive addition of 1, are distinct, and let us denote them by the usual symbols

$$1,\ 2,\ 3,\ 4,\ \cdots.$$

The class N is then *infinite*. It follows further, from the fact that corresponding to each of these elements there exists an inverse with respect to addition, that N also contains elements which, in the usual notation, are written,

$$-1,\ -2,\ -3,\ -4, \cdots.$$

Our class now satisfies the first assumption: it forms a group with respect to addition.

The second assumption requires it to form a group with respect to multiplication, except that no inverse of 0 is required. The first three assumptions for a group are satisfied, if we interpret $\times$ to be ordinary multiplication. The fourth assumption, however, requires the existence of an inverse for each of the above symbols (except 0) with respect to multiplication. This implies the existence of the "reciprocals"

$$\tfrac{1}{2},\ \tfrac{1}{3},\ \tfrac{1}{4}, \cdots \text{and} -\tfrac{1}{2},\ -\tfrac{1}{3},\ -\tfrac{1}{4}, \cdots,$$

and hence follows the existence of symbols corresponding to *all the rational numbers* But we have already seen that the latter form by themselves a number system. To characterize completely, in its abstract form, the ordinary real number system of algebra, additional assumptions are, therefore, necessary. We are in a position now to state the assumptions characterizing abstractly the ordinary system of real numbers in a very simple way.

The System of Ordinary Real Numbers. — The ordinary system of real numbers is a *class* R involving a *relation* $<$ and two *operations* $+$ and $\times$, subject to the following assumptions:

RN_1. *The class* R *is an unlimited linear continuum with respect to* $<$.

RN_2. *The class* R *is a commutative number system with respect to* $+$ *and* $\times$.

RN_3. *If* a, x, y, *are any elements of* R, *and* $x < y$, *then we have* $x + a < y + a$.

RN_4. *If* a, b, *are any two elements of* R, *and* $0 < a$, $0 < b$, *then* $0 < a \times b$.

The two assumptions RN_3 and RN_4 serve to connect the relation $<$ with the operations $+$ and $\times$ respectively. All the fundamental laws of algebra can be derived formally from these assumptions, which, moreover, may be shown to be categorical.[1] It would take us too far, however, to go into further details. A similar characterization of the system of ordinary complex numbers will be given later.

[1] This characterization of the real number system is essentially equivalent to one of the sets of postulates given by E. V. Huntington, *Transactions of the American Mathematical Society*, Vol. 4 (1903), p. 358.

LECTURE X

HISTORICAL AND LOGICAL DEVELOPMENT OF THE CONCEPT OF REAL NUMBERS

Positive Integers. Addition. Multiplication. — Historically, the first numbers to appear were doubtless the positive integers. Symbols representing them are found in the earliest records of ancient times. It happens that these numbers are precisely the ones which we have been able to define abstractly in terms of the notions of class and correspondence, since they are the finite cardinal numbers. The relation of $<$, existing between any two numbers of this kind, has also been defined. We may now define the operations of addition and multiplication for positive integers. It will be recalled that a positive integer was defined as a symbol associated with a finite class. The same symbol attached to all classes equivalent to the given one. Suppose, then, we have two classes, A and B, whose cardinal numbers are respectively a and b. Let C be a new class formed by combining the two classes A and B. In other words, let C be a class such that every element of A and of B is an element of C, and such that

every element of C is an element either of A or of B. The cardinal number c of the class C is called the *sum* of the cardinal numbers of the two classes A and B; in symbols $a + b = c$. We may then readily prove that the operation of *addition* thus defined is (1) *unique* (that is, the sum of two integers is a uniquely determined integer as soon as the two integers are given), (2) *associative*, (3) *commutative*, (4) *monotonic*. By the latter is meant that a certain relation exists between the operation of addition just defined and the relation of $<$; namely, that if two numbers m and n are in the relation $m < n$, and x is any positive integer, then $m + x < n + x$ These four properties we will call the *fundamental properties* of addition.

In order to define multiplication, let A and B, as before, be two classes whose cardinal numbers are a and b respectively. Let us form a new class C′, each element of which shall consist of a pair of elements, one taken from the class A, the other taken from the class B. The class C′ is to consist of all possible such pairs, and the cardinal number c' of this class C′ is called the *product* of the numbers a and b, in symbols $a \times b = c'$. As in the case of addition, the following four fundamental properties of multiplication are readily derived. Multiplication is (1) *unique* (that is, the product of two integers is a uniquely determined integer), (2) *associative*, (3) *commutative*, (4) *monotonic*; that is, if $m < n$, then $mx < nx$ (m, n, x being any positive integers).

Besides these eight properties of addition and multiplication, it is possible to derive from the above definition and the general properties of classes the so-called *distribu-*

tive law, which connects addition and multiplication; viz. if a, b, c are any positive integers, $a(b+c) = ab + ac$.

Rational Fractions — The first extension of the notion of number beyond that of the positive integers was made by the introduction of fractions, or, as we shall call them, the *positive rational numbers.* This extension involves a new conception of the nature of number. It is now no longer merely the result of counting, but is used to express quantity or magnitude. Symbols expressing fractions are also found in the records of very ancient times, for example, among the records of the Babylonians. Euclid considered commensurable ratios, which are now expressed by means of rational numbers, but he did not himself regard such ratios as numbers. DIOPHANTUS of Alexandria, who lived about 300 A D., seems to have been the first actually to have made use of rational numbers.[1]

From a purely logical point of view, a rational number may be regarded merely as a pair of integers. A rational number $\frac{m}{n}$ may be regarded simply as the pair of integers (m, n). It is possible then to define the relation of order and the operations of addition and multiplication for the rational numbers in terms of the relation and the opera-

[1] The Egyptian AHMES, who wrote about 1700 B C., however, made extensive use of so-called "unit fractions," *i.e.* fractions whose numerators are unity This papyrus of Ahmes is the most ancient treatise on mathematics that has come down to us Cf. CAJORI, *History of Mathematics*, p. 11. Also G A. MILLER, "The Mathematical Handbook of Ahmes," *School Science and Mathematics*, Vol. 5 (1905), p. 567.

tions already defined for positive integers. This is done briefly as follows. Two rational numbers $\frac{m}{n}$, $\frac{m'}{n'}$, are defined to be *equal*, if $mn' = nm'$. The rational number $\frac{m}{n}$ is said to be *less than* the rational number $\frac{m'}{n'}$, if $mn' < nm'$. The *sum* of the two rational numbers $\frac{m}{n}$ and $\frac{m'}{n'}$ is defined as the number $(mn' + m'n,\ nn')$. Their *product* is (mm', nn'). These definitions, it will be observed, depend merely on properties of positive integers already defined. In terms of these definitions, it is then possible to prove that the operations of addition and multiplication, as applied to the rational numbers, *satisfy the same nine fundamental laws previously noted to hold for the positive integers alone.*

A General Observation. — A general observation is perhaps in place here, which will tend to make clear our point of view in this historical survey. When a set of properties has been observed to hold for a given class of elements, and this class is then made more extensive by the addition of new objects, it may be expected that some of the properties may cease to hold. We wish to examine the successive extensions of the concept of number with reference to the nine properties which we have enumerated. The remarkable fact to be noted is what may be called *the permanence of these nine properties.* In spite of the wider and wider meaning applied to the word "number," it will be found that only two of these properties require modification.

Introduction of Irrational Numbers. — The conception of a number as expressing the measure of a quantity or magnitude soon made apparent the necessity for a further extension. Euclid, and indeed some Greek philosophers before him, were familiar with linear segments whose ratio was incommensurable. In this connection, we need only recall the ratio of the diagonal to the side of a square and of the circumference of a circle to its diameter. As has already been said, however, Euclid did not regard these ratios as numbers. The irrational numbers, as such, seem not to have appeared in western Europe until near the end of the sixteenth century A.D. It was at that time that the decimal system of notation was applied to the representation of fractions. In expressing the rational fractions in this notation, it was soon found that certain ones, for example $\frac{1}{3}$, led to nonending decimals which were "periodic." It was natural, then, for mathematicians of that age to consider also nonending decimal fractions which were not periodic. It was soon found that they led to an arithmetical expression for incommensurable ratios. They began to use these irrational numbers in the same way in which they used the rational ones, without inquiring too critically as to the justification of attributing the same fundamental properties to them. It was not until recent times that the theory of irrational numbers was placed on a scientifically satisfactory basis by WEIERSTRASS (1815–1897), DEDEKIND (1831–), MERAY (1835–), and GEORG CANTOR (1845–).

Dedekind's Postulate. "Cuts." — We will first recall Dedekind's postulate: If an ordered class C is divided into two

non-empty subclasses C_1, C_2, such that every element of C is an element either of C_1 or of C_2, and such that every element of C_1 precedes every element of C_2, then there exists an element x which effects this division. Any division of an ordered class C into subclasses C_1, C_2, as described in this postulate, Dedekind called a *cut* in the class C. Considering now the class of positive rational numbers, it is readily seen that there are two kinds of cuts; namely, cuts which satisfy Dedekind's postulate, and which we may call *closed cuts*, and cuts which do not satisfy this postulate, and which we may call *open cuts.* As an example of the first kind, we may mention the cut in which C_1 consists of all rational numbers less than 2, and C_2 consists of all rational numbers equal to or greater than 2. The number 2, in this case, defines the cut in question. As an example of an open cut, we may mention the cut already considered (p. 83), in which C_1 consists of all positive rational numbers whose square is less than 2, and in which C_2 consists of such rational numbers whose square is greater than 2. There is, then, as has been seen, no rational number which defines this cut.

Dedekind's Definition. — Dedekind's method of treating irrational numbers consisted simply in defining every cut in the system of rational numbers to be a number. The closed cuts then correspond to the rational numbers themselves; the open cuts constitute the irrational numbers It seems to be more in agreement with the popular conception of number to think of these numbers as symbols representing the cuts rather than as the cuts themselves. Assuming, then,

that a new symbol has been introduced corresponding to every open cut in the system of rational numbers, and calling every such symbol an irrational number, it is necessary to define the relation of order and the operations of addition and multiplication for these new symbols. As to order, an irrational number a defined by a cut (C_1, C_2) in the rational numbers is *defined* to lie between any number of C_1 and any number of C_2. Given two irrational numbers a and a', defined by cuts (C_1, C_2) and $(C_1' \; C_2')$, respectively, a is said to be less than a', if C_2, and C_1' have any number in common. To define the sum of a rational number m and an irrational number a defined by the cut (C_1, C_2) the number m is added to all the numbers of C_1 and to all the numbers of C_2. The two new classes C_1' and C_2', thus formed will, in view of the monotonic law for addition, constitute a new cut. The number corresponding to this new cut is defined to be the *sum* of the two numbers a and m. The sum of two irrational numbers a and b, defined by cuts (C_1, C_2) and (C_1', C_2'), respectively, is now obtained by adding, on the one hand, every number of C_1 to every number of C_1', and on the other hand, every number of C_2 to every number of C_2'. The number determined by this new cut is defined to be the *sum* of the two numbers a and b. To define the product of a rational number m and an irrational number a defined by a cut (C_1, C_2), the number m is multiplied into all of the numbers of C_1 and into all of the numbers of C_2. The cut determined by the two new classes is defined to be the *product* of m and a. Similarly, to find the product of two irrational numbers

a and b, defined by cuts (C_1, C_2) and (C_1', C_2'), respectively, a new cut is formed by multiplying, on the one hand, every number of C_1 by every number of C_1', and, on the other hand, every number of C_2 by every number of C_2'. The number corresponding to this cut is defined to be the *product* of the two numbers a and b.

The Class of Positive Real Numbers. — The nine fundamental laws for addition and multiplication may then be proved to hold for this more extended class of numbers which constitutes what is known as the class of all *positive real numbers.* The detailed discussion of these laws is, however, long. We are satisfied here to indicate its possibility. It may be readily shown thereafter that it satisfies the assumptions previously made to characterize a linear continuum.

LECTURE XI

NEGATIVE NUMBERS

Historical Items. — We have hitherto considered the extensions of the notion of number which were brought about by its introduction as a measure of quantity simply. The next extension, namely, that of introducing negative numbers, involves a radical change in the conception of number itself, and was effected only after years of struggle and controversy. The need for this extension arose first in connection with the solution of equations, and therefore did not appear until the subject of algebra was considerably developed. The first writer who appears to have recognized the existence of negative roots of a quadratic equation was the Hindu BHASKARA, in a work written about 1150 A.D. He gives $x = 50$, $x = -5$, as the roots of $x^2 - 45x = 250$; "but," says he, "the second value is in this case not to be taken, for it is inadequate; people do not approve of negative roots."[1] For centuries thereafter people did not approve of negative roots. The German mathematician, MICHAEL STIFEL, speaks, in 1544, of numbers which are "absurd," or "fictitious, below zero," and

[1] CAJORI, *A History of Mathematics*, New York, 1909, p. 93.

which arise when "real numbers above zero" are subtracted from zero. CARDAN (1501–1576) speaks of a "pure minus"; "but these ideas," says HANKEL, "occur but rarely, and until the beginning of the seventeenth century mathematicians dealt exclusively with absolute positive quantities."[1] The use of negative numbers does not appear to have been definitely established until about the time of DESCARTES (1596–1650).

Magnitude and Direction. — The reason for the great difficulty which attached to the introduction of negative numbers is doubtless due to the fact that centuries of tradition had firmly fixed in the minds of mathematicians the idea that a number must express a *magnitude* simply. The negative number adds to this notion of magnitude the notion of *direction* or *sense*. If numbers are interpreted geometrically by the points of a straight line in the familiar way by choosing an origin 0, and representing the positive numbers by points to the right of 0, the negative numbers by points on the left, it is at once seen that a number represents not merely a distance (magnitude) from 0, but also one of two directions from 0. The notion of a negative number is at present fairly well understood even outside of the mathematical public. The notion of above and below zero on the thermometer scale, the notion of credit and debit in business, etc., have served to make people in general familiar with this notion. Let us keep clearly in mind, then, that according to this new conception

[1] HANKEL, *Zur Geschichte der Mathematik im Altertum und Mittelalter*, Leipzig, 1874, p. 371; CAJORI, loc. cit., p. 152.

a number represents a magnitude in a given direction. This conception leads, as we shall presently see, to a still further extension of the notion of number. From an algebraic point of view it may be noted that negative numbers were introduced in order to make the operation of subtraction possible between any two numbers. It is interesting to observe in this connection that this very extension has made the operation of subtraction for which it was introduced unnecessary, since to subtract a number is now simply equivalent to the addition of the corresponding negative number.

Formal Treatment of Negative Numbers. — The purely formal introduction of negative numbers is as follows. If a and b are any two positive numbers, and $a < b$, the expression $a - b$ has no meaning in the system of numbers thus far considered. The expression $b - a$ in this case is, however, a definitely determined positive number c. The expression $a - b$ is then placed equal to a new symbol $-c$, which is called a *negative number*. It is then readily seen that, according to this definition, corresponding to every positive number there exists one of these new symbols which we have called negative numbers. The positive number c is called the *absolute value* of the negative number $-c$.

Order. — The order relations in this more extended class of symbols comprising the positive and the negative numbers are then given by the definition that every negative number is less than any positive number, and that of two negative numbers, $-a$ and $-b$, we have $-a < -b$, if $b < a$.

By the introduction of the new symbols the ordered class of the positive real numbers (including zero), which had a first element and no last, has been extended in the opposite direction so as to form an *unlimited linear continuum.*

Addition and Multiplication. — Addition is defined by the relations $a + (-b) = a - b$. Multiplication is defined for the new symbols by the equations $(-a)b = a(-b) = -(ab)$, $(-a)(-b) = ab$. *With these definitions,* it is possible to prove the uniqueness, the commutativity, the associativity, and the distributivity of addition and multiplication. The monotonic law for addition also holds without change. This is perhaps seen most clearly by a geometrical interpretation of addition. If the positive and negative numbers are represented in the familiar way by the points of a line, then the addition of a number x, positive or negative, to any number is represented geometrically by the movement of the whole line through a distance equal to the absolute value of x to the right or left, according as x is positive or negative. It is then apparent that, if two real numbers a and b satisfy the relation $a < b$, or geometrically that the point a is to the left of the point b, the same relation will hold between the resulting points when the line has thus been shifted through any distance to the right or left.

The monotonic law for multiplication, however, now requires modification. It will be remembered that this law, in its original form, stated that if $a < b$, then $ax < bx$. We have seen that the law held for all positive numbers, a, b, x. It is at once seen, however, that it no longer holds when x is a negative number or zero. Indeed, while 2 is less than

3, we know that 2 (-1) is not less than 3 (-1). The law must then be modified to read as follows: *If* $a < b$, *then* $ax \lesseqqgtr bx$, *according as* $x \gtreqqless 0$. We have here the first example in which an extension of the number system made necessary a modification of one of the nine fundamental laws of operation previously enumerated.

Law of Signs cannot be Proved. — The rules for addition and multiplication of negative numbers were given above as definitions. One often sees attempts to prove these rules. That this is a logical impossibility without some additional definition or assumption follows immediately from the fact that other rules for addition and multiplication may be given which are quite as logically consistent with the previous laws as those which are familiar. These rules must therefore be regarded as *pure conventions* concerning the use of the new symbols introduced into the system of numbers. If it is asked, whence come these conventions, we will find the answer in a general principle which has already been referred to incidentally, and which may be called the *permanence of the formal laws.*

The conventions as to signs are made so as to preserve this permanence purely on the ground of the convenience and serviceability of the resulting symbolism.

Negative Numbers forced their Introduction. — We have seen that negative numbers came into use only gradually during the development of algebra, and that their adoption was practically forced upon the mathematicians of the sixteenth, seventeenth, and eighteenth centuries, without

any conscious volition on their part, and, in fact, in spite of much active opposition. It really looks, to quote from Professor Klein, as though the algebraic symbols were more reasonable than the men who employed them. The rule of multiplication, expressed by the equation

$$(1) \qquad (a-b)(c-d) = ac - ad - bc + bd,$$

was known and used in the case where $a-b$ and $c-d$ were positive numbers from the very beginning of the development of algebraic methods. The rule in this case has a very simple geometric interpretation of the kind which older mathematicians delighted in using as proofs. In the attached figure (Fig. 11), the numbers a, b, $a-b$ are represented by segments of a horizontal line. Similarly the numbers c, d, $c-d$ are segments of a vertical line. The product of the two numbers $a-b$ and $c-d$ is then represented by the area of the unshaded rectangle in the figure, and it is readily seen that this area may be obtained from the

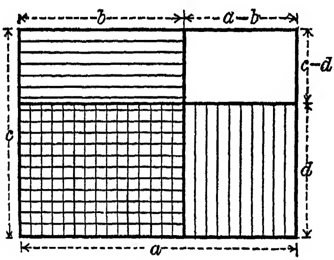

FIG. 11

whole rectangle, which is represented by ac, by subtracting successively the areas bc and ad of the horizontally and vertically shaded rectangles respectively, provided the rectangle bd (which has thus been subtracted once too often) is again added. The equation (1) is one of the formal laws

referred to. If it is *assumed* to hold for all values of a, b, c, d, we find in particular, for the values $a=0$ and $c=0$, that it gives $(-b)(-d)=bd$. It is the fact that the arbitrary definition of addition and multiplication for the negative numbers preserves such formal laws that makes these particular conventions so serviceable. Let it be emphasized again that there can be no such thing as an *a priori* proof of these laws of signs, but that they are pure conventions, finding their justification on the logical side in their consistency with previous assumptions and on the practical side in their serviceableness.

Consistency of Operations with Negative Numbers. — "But how do we know," it will be asked, "that the definitions thus made are logically consistent?" This question did not receive an answer until the nineteenth century, and here, as in all cases, no absolute proof of consistency is known. The most that we can do is to reduce the question of consistency to that of the system of ordinary positive integers, or to give a concrete representation of the symbols and of the relations and operations involved, which appears to satisfy all the assumptions made. For the former method, *i.e.* that of reducing the consistency of the algebra of negative numbers to that of the system of positive integers, we will refer to the treatment given in PIERPONT'S *Theory of Functions of a Real Variable.* We will, however, give a concrete geometric interpretation of the operations involved which will be of use to us on a later occasion. Let the system of all real numbers (including the negative) be represented as before by the points of a line. The opera-

tion of addition has already been interpreted. As to the interpretation of multiplication, we may note that to multiply any number by any positive number is equivalent geometrically to an expansion away from or a contraction toward the origin of the segments of the line of representation, according as the multiplier is greater than or less than one. If the multiplier is a negative number, the geometric interpretation consists likewise of such an expansion or contraction, determined by the absolute value of the multiplier, combined with a rotation of the whole line about the origin through an angle of 180° (Fig. 12). Since there is

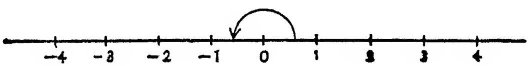

FIG. 12

nothing self-contradictory in the conception of these displacements and expansions or contractions, there can be nothing self-contradictory in the operations with negative numbers to which they correspond.

Two Characterizations of the System of Real Numbers. — The real number system of algebra has now been characterized by two methods. The first, given at the end of Lecture IX, described this number system at one stroke, as it were, by a set of assumptions concerning the undefined terms, *number*, $<$, $+$, $\times$. The other, the discussion of which we have just completed, followed a more historical line of development, and showed how it was possible to build up this system by successive extensions of the number concept, beginning with that of the positive integer. It could now be

shown without serious difficulty that the two characterizations of the real number system which have been given are abstractly equivalent.

Bearing on Elementary Teaching. — It will be well now to ask ourselves what bearing this discussion has on the problems of elementary teaching. Should either of these methods be applied in making the pupil, at the first approach to the study of algebra, familiar with its conceptions and its processes? Should we begin by placing before him a set of assumptions characterizing real algebra, or should we attempt to make clear to him the notions involved by successively extending the number system in the way described above? It will probably be granted, without argument, that neither of these methods is suitable for the purpose in hand. As far as the real positive numbers are concerned, little, if anything, need be added in a first course in algebra to the knowledge the pupil already possesses of these numbers from his study of arithmetic. The concrete representation of these numbers as magnitudes, in particular the geometrical representation by means of distances measured along a line, should, of course, be emphasized and used to clarify the pupil's conception of the nature and the use of these numbers.

The first great difficulty arises in the introduction of negative numbers. The reason for this difficulty is found in the significant fact that, with this introduction, algebra changes its aspect. *It can no longer be regarded as a system of concrete numerical magnitudes, but assumes the aspect of a formal symbolism.* This being its true nature, the pupil

should gradually be led so to regard it. At the beginning, the change of attitude which is required of him, and which demands to a very high degree the power of abstraction, should be made as easy as possible by the frequent use of geometric and other representations, and especially by the continued application of the symbolism to concrete problems, which will slowly, but surely, convince the pupil of its great usefulness.

Here, as in geometry, the fundamental mistake is continually made of trying to prove everything. The commutative, associative, and distributive laws should be presented to him as properties with which he is already more or less familiar from his study of arithmetic, although he may not previously have stated them explicitly. Certainly no attempt should be made to prove them. Equally absurd, both scientifically and pedagogically, is any attempt to prove the fundamental laws of signs "minus times plus equals minus, minus times minus equals plus." These laws should rather be presented to him as "rules of the game," concerning the justification of which he will become increasingly convinced as his study proceeds.

It may be well to call particular attention at this point to one method of "deriving" the law of signs, because it has found its way into so many of our current texts. This method consists of defining the product $a \times b$ by saying that it is equal to the result of performing upon a those operations which must be performed upon unity to get b. The product $3(-2)$, for example, is then shown to be equal to -6, by observing that -2 is obtained from unity by

changing 1 to -1 and adding the result to itself. According to this definition, in order to obtain $3(-2)$, we would then change 3 to -3 and add it to itself, -3 plus -3 giving -6. This method could in any case be used only for the derivation of the product of particular numbers; there is no way of deriving from it the general formula that $a(-b) = -(ab)$. But aside from this, from the logical point of view, it is entirely worthless, since the rule implies that every number which is used as a multiplier may be obtained from 1 by a definite series of operations, whereas it is obvious that the multiplier may be obtained from 1 in many different ways, and the proof to be complete would require a demonstration that all these ways of deriving the multiplier from 1 lead to the same resulting product.

It will be seen, then, that our teaching should relate itself more closely to the first method given for characterizing the system of real numbers than to the second, in view of the fact that no attempt should be made to give formal definitions of number or of the operations of multiplication and addition, and that the laws governing the latter should be assumed without proof. On the other hand, elements of the second method are also present, in that the number system is to be extended to include the negative numbers, and later the imaginary and complex numbers, only after the student is familiar with the algebraic processes applied to the positive numbers.

Classification of Real Numbers. — Before closing the discussion of the system of real numbers, it seems desirable to

say a word regarding the classification of these numbers. We have already spoken of the distinction between rational and irrational numbers. The former are defined as those numbers which can be expressed as the quotient of two integers. They include, according to this definition, the integers themselves. This definition is so simple, it is strange how much confusion exists regarding it. Ninety-nine students out of a hundred, if asked to define a rational number, will attempt to give a negative definition, to the effect that a rational number is one which does not "contain radicals," a definition which is obviously inadequate.

We are concerned here primarily with another classification, viz. into so-called *algebraic* and *transcendental* real numbers. An algebraic real number is defined as any real number which is the root of an algebraic equation,

$$a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0,$$

in which the coefficients, a_0, a_1, $\cdots$, a_n, are integers, positive, negative, or zero. Any real number which is not algebraic is said to be a transcendental (real) number.

The Algebraic Numbers are Denumerable. — The existence of such transcendental numbers follows at once from the following consideration. It may be readily proved that *the class of all algebraic real numbers is denumerable.* We will assume the truth of the following well known theorem from the theory of equations to the effect that every algebraic number is a root of one and only one so-called irreducible equation, *i.e.* one the left-hand member of which cannot be resolved into factors of the same form and of lower degree,

with integral coefficients, and in which the coefficients, a_0, a_1, $\cdots$, a_n, have no common factors greater than unity. CANTOR effected the proof of the theorem in question by attaching to every such irreducible equation a positive number,

$$N = n - 1 + |a_0| + |a_1| + \cdots + |a_n|,$$

which he called the *height* of the equation.[1] It is then easily seen that the number of irreducible equations of given height is finite. All algebraic numbers may then be put into one-to-one correspondence with the positive integers by considering first all equations of height 1, then all equations of height 2, then all equations of height 3, and so on, and arranging the roots of all possible equations of a given height in order of magnitude. Every such root, in other words every algebraic number, is thus assigned a definite place in the progression of positive integers.

Recalling the fact that the class of all real numbers is non-denumerable, it follows from the theorem just proved that the class of all algebraic numbers does not comprise all the real numbers. In fact, it shows that, compared with the class of all real numbers, the algebraic numbers form an almost insignificantly small part. In spite of this fact, very little is known regarding transcendental numbers as such. It is, in general, a very difficult problem to determine of a given real number whether it is algebraic or transcendental. The first number to be proved transcendental was the base of the natural system of logarithms, usually denoted by the number e. This proof was given by the

[1] The notation $|a_i|$ means "the absolute value of a_i."

French mathematician Hermite, in 1874. Not until the year 1882 was the attempt to prove the number π transcendental successful. The proof was published in that year by a German, Lindemann. It is beyond the scope of these lectures to give either of these proofs. Regarding the bearing of the transcendental character of π on the famous problem of the squaring of the circle, we shall have something to say in a later lecture.

LECTURE XII

ORDINARY AND HIGHER COMPLEX NUMBERS

Introduction of Complex Numbers. — The so-called imaginary or complex numbers forced their way into algebra in a way very similar to that of the negative numbers. Complex numbers occurred apparently as early as 1545, in CARDAN's solution of the cubic equation, but their occurrence at that point is only incidental; and, as in the case of negative numbers, it took centuries before the scruples of mathematicians against the use of such numbers were overcome. Until the end of the eighteenth century, these numbers seem to have had a rather mysterious aspect, just as to-day every pupil is inclined to be mystified when first he hears of that peculiar $i = \sqrt{-1}$.[1] Not until the early part of the nineteenth century was the real nature of these numbers recognized, and their use placed upon a strictly logical foundation. Just as negative numbers became necessary in order to make the operation of subtraction always possible, so the introduction of complex numbers became

[1] A number of important theorems involving complex numbers date from this time, however. We may mention in this connection the work of DE MOIVRE (1667–1754) and of EULER (1707–1783).

necessary in order to make the extraction of roots always possible. From the point of view developed in the last lecture, in connection with the introduction of negative numbers, whereby algebra is to be regarded as a formal symbolism which is to be studied on account of its useful applications, there should be little difficulty in recognizing that a still further extension of this symbolism might be possible and useful. Furthermore, as will be seen presently, the usual interpretation of a complex number involves little that is essentially different from the conception which we have already attached to a negative number, namely, that of representing the combination of a magnitude and a direction.

Consistency. — We should be concerned here primarily with the question as to the logical consistency of the new symbolism. That the algebra of the ordinary complex numbers $x + iy$, where x and y are real numbers, is logically consistent is seen very easily as follows. From an abstract point of view, the number $x + iy$ may be regarded simply as a pair (x, y) of real numbers, x and y. Two complex numbers, $x + iy$ and $x' + iy'$, or the two corresponding pairs (x, y) and (x', y'), are defined to be equal if and only if $x = x'$ and $y = y'$. The sum of two numbers (x, y) and (x', y') is defined as the complex number, or number pair, $(x + x', y + y')$. With these definitions, the fundamental laws for addition, with the exception of the monotonic law, are readily seen to hold. The monotonic law, of course, breaks down, in view of the fact that the system of complex numbers no longer forms a linearly ordered class. As for

multiplication, the product of two complex numbers (x, y) and (x', y') is defined to be the complex number $(xx' - yy', xy' + yx')$; and, with this definition, the fundamental laws for multiplication, except the monotonic law, which breaks down for the same reason as noted for the case of addition, are readily proved to hold. The distributive law of multiplication with respect to addition also follows without difficulty. These considerations show that there can be nothing inconsistent in this new symbolism, since any such inconsistency would involve a contradiction in the system of real numbers, *i.e.* in the pairs of real numbers which we have been considering. It follows from these definitions that algebraic operations on complex numbers $x + iy$ may be performed according to the laws established for real numbers, except that the symbol i^2 is to be replaced, wherever it occurs, by the number -1.

A Geometric Interpretation. — A geometric interpretation of complex numbers may be obtained as follows. We have seen in a previous lecture that when the real numbers are represented by the points of a line the operation of multiplying every number of this system by -1 is equivalent to the rotation of the line about the origin in a plane through an angle of 180°. According to what has been said, the operation of multiplying any real number by i is such that when performed twice the result must be equivalent to the multiplication of the number in question by -1. The natural geometric interpretation of multiplying all the real numbers by i is, then, that this multiplication is equivalent to a rotation of the line, the points of which

represent the real numbers, and which we will call the *axis of reals,* about the origin in a plane through an angle of 90°. This operation evidently satisfies the condition that when repeated it gives an operation equivalent to multiplying by -1 (Fig. 13). The so-called pure imaginary numbers, that is, those of the form iy, where y is a real number, are then represented by the points of a line, called the *axis of imaginaries,* passing through the origin and perpendicular to the axis of reals. Any other complex number, $x+iy$, is then represented by a point, in the plane of these two lines, whose coördinates with respect to the two axes in question are x and y, respectively, the axis of reals playing the rôle of the axis of abscissas and the axis of imaginaries that of ordinates. By this means, every complex number corresponds to a unique point of the plane, and, conversely, every point in the plane represents a unique complex number. Such a representation would, however, be of little value, if it were not possible to interpret conveniently the fundamental operations of addition and multiplication. Let us note first that a complex number may also be

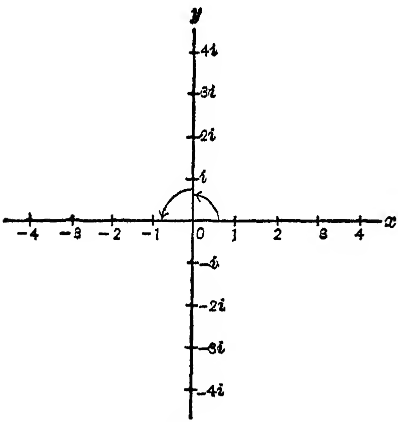

Fig. 13

thought of as representing not a point in the plane, but the line joining the origin to the point in question; that is, by a line of a certain length issuing from the origin in a certain direction, in other words, by a *vector*. We see here what was meant when it was stated that the ordinary interpretation of complex numbers conceives of such a number as representing a magnitude combined with a direction. It is distinguished from the real numbers, from this point of view, merely in the fact that, while the latter could repre-

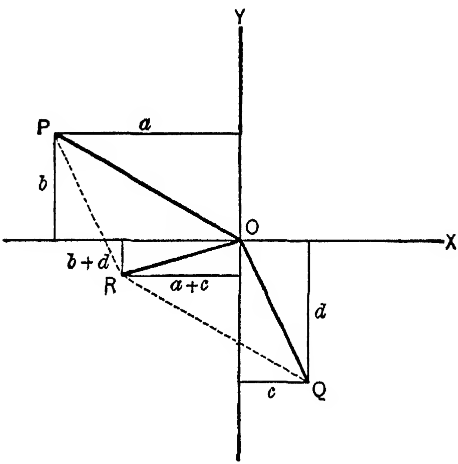

FIG. 14

sent magnitudes combined with one of two opposite directions only, the complex numbers represent magnitudes combined with any direction in a plane.

Addition and Multiplication — Thinking of the complex number as represented by a vector issuing from the origin, we

may represent the sum of the two complex numbers $a+ib$ and $c+id$, corresponding respectively to the vectors OP and OQ (Fig. 14), in magnitude and direction by the diagonal OR of the parallelogram of which OP and OQ are two adjacent sides. In order to describe simply the geometric interpretation of multiplication, let us first define what is meant by the absolute value and the argument of a complex number. The *absolute value* of a complex number $x+iy$ is defined as the positive value of $\sqrt{x^2+y^2}$. It evidently represents geometrically the length of the vector representing the number. The angle which this vector makes with the positive end of the axis of reals is called the *argument* of the complex number. The product of the two numbers $a+ib$ and $c+id$ is then readily shown to be a number whose absolute value is the product of the absolute values of the two vectors, and whose argument is the sum of the arguments of the two vectors. It is represented geometrically, therefore, by a vector OS, whose length represents the product of the two absolute values and which makes an angle with the axis of reals equal to the sum of the two angles made by the vectors OP and OQ This geometrical interpretation provides another proof of the logical consistency of the algebra of ordinary complex numbers.[1]

Generalizations. — Having conceived of algebra as a formal symbolism with useful concrete interpretations, it is per-

[1] This geometric interpretation of complex numbers has until recently been attributed to GAUSS (1799) and ARGAND (1806) It was anticipated, however, by CASPAR WESSEL in a paper presented to the Danish Academy of Sciences in 1797.

fectly natural for us to suppose that there may be other symbolisms which may also prove serviceable. As far as the domain of ordinary algebra is concerned, no further extension of the system of complex numbers is necessary, that is to say, all algebraic operations performed upon numbers of this system are now possible and lead to numbers of the same system. In fact, it can be shown that no further extension of this number system is possible without the sacrifice of one or more of the fundamental laws of algebra. Among the latter are to be included now *the uniqueness of the operations of subtraction and division, and the fact that the product of two numbers cannot be zero unless one of the factors is zero.*

Vector Analysis. — The geometric interpretation of complex numbers at once suggests a possible symbolism that may be of value. We have seen that complex numbers may be represented by vectors in a plane. If we consider the important applications which the notion of vector has in physics, especially in mechanics, — we need only recall that velocities, accelerations, forces, etc., are represented by vectors, — it is to be expected that a symbolism the elements of which represent *vectors in space* may be of the greatest value, and such indeed is the case. The so-called *vector analysis* has of recent years reached a high degree of development, and is being used more and more extensively in the mathematical treatment of physical problems.

Higher Complex Numbers. — On the other hand, the form of complex numbers, which may be thought of as consisting of all possible combinations obtained by multiplying two

units, 1 and i, by any pair of real numbers, x and y respectively, and adding the results, viz. $1 \cdot x + i \cdot y$, at once suggests a possible generalization of which the so-called *quaternions* are perhaps the most important example. These two methods of generalization, leading to vector analysis and the algebra of quaternions, one resting on a geometric, the other on a purely formal foundation, were developed almost simultaneously by the German, H. GRASSMANN, and the Englishman, W. R. HAMILTON, respectively, about 1840. We ought, perhaps, to give a brief account of the algebra of quaternions, but first it may be well to consider the general formulation suggested above.

In place of the two "units" 1 and i, we suppose given any finite number n of "units," which we will denote by e_1, e_2, $\cdots$, e_n; and form a "number"

$$a = a_1e_1 + a_2e_2 + \cdots + a_ne_n,$$

the coefficients a_1, a_2, $\cdots$, a_n being any ordinary real numbers.[1] Two such numbers, a and

$$b = b_1e_1 + b_2e_2 + \cdots + b_ne_n,$$

are defined to be equal if, and only if, the corresponding coefficients throughout are equal; that is, if

$$a_1 = b_1,\ a_2 = b_2,\ \cdots,\ a_n = b_n.$$

The sum and difference of any two such numbers is simply defined by the expression

$$a \pm b = (a_1 \pm b_1)e_1 + (a_2 \pm b_2)e_2 + \cdots + (a_n \pm b_n)e_n.$$

[1] The following discussion follows substantially the treatment by KLEIN, *Elementarmathematik vom höheren Standpunkte aus*, vol. I, pp. 144 *et seq.*

The commutative and associative laws of addition are then readily seen to hold for this system of numbers.

Multiplication of Higher Complex Numbers. — The problem becomes more complicated and more interesting as soon as we attempt to define multiplication. It is natural to proceed at the outset in accordance with the ordinary distributive law of algebra and place the product

$$a \cdot b = (a_1 e_1 + a_2 e_2 + \cdots + a_n e_n)(b_1 e_1 + b_2 e_2 + \cdots + b_n e_n),$$

which, when multiplied in the ordinary way according to the *distributive law*, will give the expression

$$a \cdot b = \Sigma_{ik} a_i b_k e_i e_k, \quad (i, k = 1, 2, 3, \cdots, n).$$

In order that this new expression shall be a number of our original system, however, it is necessary that the products of the units e_i, e_k, shall be numbers of the system, that is, must be linear combinations of the original units $e_1, \cdots, e_n$. In order to define multiplication, then, we must have n^2 equations of the form

$$e_i e_k = \Sigma_l c_{ikl} e_l, \quad (i, k, l = 1, 2, 3, \cdots, n),$$

where the c_{ikl} are real numbers. Substituting these equations in the expression for the product obtained above, we have

$$a \cdot b = \Sigma_l \Sigma_{ik} a_i b_k c_{ikl} e_l, \quad (i, k, l = 1, 2, 3, \cdots, n),$$

which is indeed a number of the form desired. In the arbitrary choice of this rule of multiplication, that is, of the system of coefficients c_{ikl}, we find the characteristic property of any particular system of these so-called *higher complex numbers*. If now we define division as the inverse of multiplication, that is, if we attempt to determine the number

b by the equation $a \cdot b = c$, a and c being given numbers of our system, we will find that the coefficients $b_1, b_2, \cdots, b_n$ of the number b are determined by n linear equations. Now it can well happen with an arbitrary choice of the coefficients c_{ikl} that this system of equations may have no solutions, or an infinite number. It is also possible that the product of two numbers a and b may be zero without either of the factors being zero. To avoid such difficulties, it is necessary to subject the coefficients c_{ikl} to further conditions.

Commutative Law for Multiplication usually Sacrificed. — But the fact should here be noted without proof that, if the number of units is greater than two, it is not possible to choose such a system of coefficients without doing violence to one or more of the ordinary laws of algebra. The law which it has usually been found most convenient to sacrifice in this case is the commutative law for multiplication.

Quaternions. — With this statement of the general problem in mind, we may now turn our attention to a brief description of the system of quaternions. As the name implies, it is a system formed by means of four units, one of which is taken to be the real number one, and the other three of which, to adopt Hamilton's notation, are denoted respectively by i, j, k. Any quaternion q is then of the form

$$q = a + bi + cj + dk.$$

The sum of two quaternions, q and q' ($q' = a' + b'i + c'j + d'k$), is then given by the expression

$$q + q' = (a + a') + (b + b')\,i + (c + c')\,j + (d + d')k.$$

As to multiplication, we must, as has been seen, first define

what shall be meant by the product of the units. Hamilton defined these by the following $4^2=16$ equations:

$$1^2=1, \qquad i\cdot 1=1\cdot i=i,$$
$$k\cdot 1=1\cdot k=k, \qquad j\cdot 1=1\cdot j=j.$$

For the squares of the units i, j, k, we place

$$i^2=j^2=k^2=-1,$$

and finally complete the list of sixteen defining equations by placing

$$j\cdot k=i, \qquad k\cdot i=j, \qquad i\cdot j=k;$$
$$k\cdot j=-i, \qquad i\cdot k=-j, \qquad j\cdot i=-k.$$

We note at once that *the commutative law of multiplication has been sacrificed by this definition.* With these rules for multiplying units, and making use of the ordinary distributive law of multiplication, the product of any two quaternions q and q' may be readily formed, and is found to be another quaternion. The associative law for multiplication, as well as for addition, is readily proved to hold. The fact which above all others merits attention in this connection is that, with this definition of multiplication, division continues to be unique. That is, given any two quaternions, q and q', $q\neq 0$, there is one and only one quaternion x satisfying the relation $qx=q'$, and the equation $qq'=0$ cannot be satisfied unless either q or q' is itself 0. It would take us too far to give the proof of these statements. They involve nothing, however, but simple algebraic computation. It would take us too far also to discuss the geometric and other applications of this system of higher complex numbers. Suffice it to say that it bears a very close rela-

tion to vector analysis already mentioned, and that it has found many valuable applications in mathematical physics.

General Concept of a Number System. — The system of quaternions has been introduced at this point merely in order to exhibit one example of a symbolism, or, as we may say more precisely, of a number system different from that of ordinary algebra, and, in particular, in which the commutative law of multiplication does not hold. Many such number systems have been discussed and have proved of greater or less value in various mathematical and physical problems. It is in connection with the consideration of all such systems that the general notion of a number system was evolved. Mathematicians have asked themselves, "what properties shall we insist upon in a symbolism in order that we may call it a number system?" The answer to this question was formulated in a preceding lecture, when the definition of such a general number system was given. To recall it briefly, it was to consist of a class of elements (numbers) in which two operations, called respectively addition and multiplication, are defined and operate subject to the conditions that the class shall form a group with respect to addition and a group with respect to multiplication, except that no inverse of zero shall be required, and that the distributive law shall hold between the two operations of addition and multiplication. From this point of view, ordinary algebra, whether real or complex, should be regarded simply as one example out of many of a symbolism satisfying this definition.

A Set of Assumptions for Complex Algebra. — In closing

this lecture, it may be desirable to give a set of assumptions characterizing the system of ordinary complex numbers:[1]

A class K of elements (numbers) is said to form a system of ordinary *complex numbers* provided it contains an undefined sub-class C and order relation $<$, and two operations $+$ and $\times$ exist and operate upon the elements of K subject to the following assumptions:

CN_1. *The class* K *forms a commutative number system with respect to the operations* $+$ *and* $\times$.

CN_2. *The sub-class* C *forms a system of ordinary real numbers with respect to* $+$, $\times$, *and* $<$.

CN_3. *The class* K *contains an element* i *such that* $ii = -1$.

CN_4. *If* a *is any element in* K, *there are two elements,* x *and* y, *of* C, *such that* $a = x + iy$.

[1] The set of assumptions which follows is substantially that given by E. V. HUNTINGTON, "A Set of Postulates for Ordinary Complex Algebra," *Transactions of the American Mathematical Society*, vol. VI, 1905. Reference may also be made to two other articles by Professor HUNTINGTON: "The Fundamental Laws of Addition and Multiplication in Elementary Algebra," *Annals of Mathematics*, second series, vol. 8 (1906), p. 1; and "The Fundamental Propositions of Algebra," *Mathematical Monographs*, edited by J. W. A. Young, Longmans, Green and Co., 1911.

LECTURE XIII

GEOMETRY. HILBERT'S ASSUMPTIONS

Introduction. — We shall now leave for the time being the subject of algebra, and turn our attention to the foundations of geometry. We have seen in the first few lectures what the problem is that confronts us. It is desired to choose a set of undefined terms and a set of unproved propositions (assumptions, postulates, or axioms) concerning them, with the property that the whole content of the so-called elementary euclidean geometry may be derived from them by the methods of formal logic, that is, without any further appeal to intuition. In stating the problem in this way, we purposely leave aside the discussion of the psychological and philosophical questions pertaining to the genesis of our space conceptions. Those who are interested in these questions may be referred to the discussion of them given by POINCARÉ in his *Science and Hypothesis,* to which we have already had occasion to refer, and by E. MACH, *Space and Geometry,* an English translation of which, by T. J. McCORMACK, has been published by the Open Court Publishing Co., Chicago. It may be noted in passing that psychologists, philosophers, and mathematicians who have

discussed these questions are by no means agreed We are concerned here only with the purely logical side of the problem.

Projective and Metric Geometry. — That the problem of setting up such a set of fundamental assumptions characterizing our ordinary geometry is very complex, will become clear as soon as we consider the large number of terms which are in use and the resulting immense freedom of choice that exists in the selection of those that are to be left undefined. We need think only of the terms "point," "line," "plane," "segment," "length," "distance," "angle," "congruent," "movement," "between," etc. Although the type of geometry we are to consider is usually referred to by the word "elementary," it is by no means the most elementary, geometry. It is complicated by the fact that it is concerned not merely with the relative positions of points, lines, planes, etc., but that it combines with such notions the ideas of magnitude and of measurement. It is by virtue of the latter fact, that the geometry which we are discussing is also referred to as metric geometry, more precisely, as *euclidean metric geometry*, a term which we shall adopt in the future. It serves to distinguish this geometry from the so-called *projective geometry* or the *geometry of position*, as it is sometimes called, which is entirely independent of the notion of measurement, and involves only the various intersectional properties of points, lines, planes, etc. The simplest method of building up a set of assumptions for euclidean metric geometry would be to study first the foundations of this projective geometry, then to add

further assumptions to characterize the metric. However desirable this method of procedure would be, it would carry us too far to follow it in these lectures. It seems desirable, however, on account of the great importance of projective geometry, to give at least one theorem which may serve to illustrate the difference between the theorems of projective and those of metric geometry. The theorem we choose for this purpose is the following:

The Theorem of Desargues. — *If two triangles ABC and $A'B'C'$ are so related that the lines AA', BB', CC' all pass through the same point O, then the pairs of corresponding sides, that is, AB and $A'B'$, AC and $A'C'$, BC and $B'C'$, intersect in three points which are on the same straight line, provided none of these corresponding sides are parallel.*[1]

Projection and Section. — This theorem is evident if the two triangles are in different planes. It can readily be shown to hold, however, when the two triangles are in the same plane. The characteristic property of the theorems of projective geometry, as distinguished from those of

[1] The condition last stated is necessary in the statement of this theorem as long as we make use of the conceptions of metric geometry. From the point of view of projective geometry, the theorem is always true, even without this restriction. This is made possible by attributing to two parallel lines an "ideal point of intersection," called "a point at infinity," and likewise to two parallel planes, a line of intersection, called "a line at infinity," etc. The introduction of these ideal elements involves merely a change in the terminology employed, and in no way affects the meaning of the proposition. For a full discussion, reference may be made to Reye, *Geometry of Position*, English translation by T. F. Holgate, New York, 1898; or Veblen and Young, *Projective Geometry*, vol. I, Boston, 1910, p. 7.

metric geometry, lies in the fact that they state properties of figures which remain valid when the figures in question are subjected to any process of *projection and section.* The following considerations will illustrate what is meant. Suppose we think of two triangles in the same plane which satisfy the conditions of the theorem just stated. This figure changes its aspect if we look at it from different angles. These different aspects of the figure are obtained geometrically by the following process. When one looks at this figure, the rays of light pass from its points and lines to the eye, O, of the observer. Corresponding to every point in the figure there is one line through O; corresponding to every line in the figure, there is a plane through O. Now the "picture" which the observer gets of this figure is obtained by taking a plane section of this system of lines and planes through O. It is readily seen that corresponding to every point of the original figure there would be a new point in the plane of section, and to every line there would be a new line. Moreover, and this is the important fact, the points and lines in the new plane will again form two triangles related in such a way that lines joining certain pairs of vertices of the two triangles all meet in a point, and that the pairs of corresponding sides of the two triangles intersect in points that are collinear. The new figure is said to have been obtained from the original one by a process of projection and section, or by a *projective transformation.* We have seen that the theorem given states a property of a plane figure which is left unchanged when this figure is subjected to a projective

transformation. This is the characteristic property of the theorems of projective geometry. It is at once evident that properties involving the notion of measurement can have no place in projective geometry.

Geometry Characterized by a Group of Transformations. — We may, at this point, answer a question of the greatest importance which at once suggests itself: What is it that characterizes the theorems of metric geometry as such? Briefly this: *The theorems of metric geometry state properties of figures which remain valid when the figure is subjected to any rigid movement or displacement in space.* This is merely another kind of transformation. So we see here illustrated one of the fundamental modern conceptions of geometry — that *a branch of geometry may be characterized by a group of transformations.* The theorems of any such branch of geometry state all those properties of figures which remain valid when the figures are subjected to any one of the transformations of the corresponding group.[1] We shall see presently what a fundamental rôle the idea of rigid movements or displacements, which we have just referred to, may be made to play in the foundations of geometry.

Hilbert's Assumptions. — To return to the problem in hand, namely, that of characterizing by means of a set of assumptions the euclidean metric geometry, we must content ourselves with a brief outline of two of the more

[1] This conception was first formulated by F. Klein, *Vergleichende Betrachtungen über neuere geometrische Forschungen*, Erlangen, 1872; reprinted in *Mathematische Annalen*, vol. 43 (1893), p. 63. English translation by M. W. Haskell, *Bulletin of the American Mathematical Society*, vol. 2 (1893), p. 215.

important methods which have been employed for this purpose. The amount of work which has been done on this subject in the past thirty years is very large. To mention only the most important names in this connection, we should cite PASCH, HILBERT, F. SCHUR in Germany, PEANO and PIERI in Italy, E. H. MOORE and O. VEBLEN in this country. The sets of assumptions which appear to us most closely related to the needs of elementary instruction are those of HILBERT and PIERI. We shall proceed to describe them as briefly as possible. Hilbert[1] considers a class of undefined elements, called *points*, and certain undefined sub-classes[2] of these points, called *straight lines* and *planes*. He divides his set of assumptions into five sub-sets: I. Assumptions of alignment; II. Assumptions of order; III. Assumptions of congruence; IV. A parallel assumption; V. Assumptions of continuity.

The Assumptions of Alignment. — The first set of assumptions, that is, the assumptions of alignment, comprise the following eight:

I. 1 *and* I. 2. *Two distinct points A and B determine one, and only one, straight line.*

I. 3. *A straight line contains at least two points; a plane*

[1] *Grundlagen der Geometrie*, first edition 1899, third edition 1909, English translation by E. J. TOWNSEND, *The Foundations of Geometry*, Open Court Publishing Co., 1902.

[2] Hilbert does not state that the "lines" and "planes" are to be thought of as classes of points. To think of them as such, however, simplifies to some extent the discussion without essentially altering Hilbert's point of view.

contains at least three points which do not belong to the same straight line.

I. 4 *and* I. 5. *Any three points which do not belong to the same straight line determine one, and only one, plane.*

I. 6. *If two points A and B of a straight line a are in a plane α, then every point of a is a point of α.*

I. 7. *If two planes α and β have a point A in common, they have at least one other point B in common.*

I. 8. *There exist at least four points which do not all belong to the same plane.*

Among the theorems derivable from this set of assumptions I. 1 to I. 8, we may mention the following: Two straight lines in a plane have either one or no point in common; two planes have either no point or a straight line in common; a plane and a straight line not in the plane have one or no point in common; there exists one, and only one, plane containing a given straight line and a point not on that line. I. 7 and I. 8, moreover, imply the so-called three dimensionality of space. The subject of dimensions is one to which we shall return later.

The Assumptions of Order. — The second set of assumptions, viz., the assumptions of order, characterize an undefined relation existing between the points of a straight line, which is expressed by saying that a point B is "*between*" two other points A and C. These assumptions were first stated by PASCH,[1] a fact to which we already had occasion to refer in Lecture IV. They are as follows:

[1] *Vorlesungen über neuere Geometrie*, Leipzig, 1882.

II. 1. *If A, B, and C are points of a straight line and B is between A and C, then B is also between C and A.*

II. 2. *If A and C are two points of a straight line, there exists at least one point B which is between A and C, and at least one point D such that C is between A and D.*

II. 3. *Of any three points of a straight line, one and only one is between the other two.*

These assumptions imply that the points of a straight line form a linearly ordered dense class in the sense in which we defined the terms in Lecture VII. At that time, we took the relation "precedes" as the fundamental one, and were able to define the term "between" in terms of it. It is very easy conversely to show from the above assumptions that, if the term "between" is taken as fundamental, the relation of precedence may be defined in terms of it. The assumptions of linear order which we gave previously then follow as theorems from assumptions II. 1, II. 2, and II. 3, given above There is one more assumption of order in Hilbert's set, the statement of which requires the following definition: Given two points, A and B, of a straight line, the class of points consisting of A and B, and all the points between A and B, is called the *interval* determined by A and B. The points A and B are called the *extremities* of the interval; the points between A and B, the *interior* points of the interval; all points of the straight line which are not points of the interval AB are said to be *exterior* to AB. The assumption in question may now be stated as follows:

II. 4. *Given three points, A, B, C, which are not on the same straight line, and a straight line a in the plane ABC, not containing any of the points A, B, or C; if the straight line a contains a point of the interval AB, then it contains also a point either of the interval BC or of the interval AC.*

As consequences of the two sets of assumptions I and II, we will merely mention the following: Between any two points of a straight line there is an unlimited number of points; given any four points of a straight line, it is always possible to denote them by letters A, B, C, D, in such a way that B is between A and C, and also between A and D, and that C is between A and D and also between B and D;[1] any straight line a of a plane α divides the points of α which are not of a into two regions with the following property: Every point A of one region determines with every point B of the other region an interval AB containing a point of a; on the other hand, any two points A and A' of the same region determine an interval AA' not having a point in common with a. It is possible by means of this theorem to define what is meant by the *sides* of a straight line in a plane. A similar theorem states that any plane α divides all remaining points into two regions with a similar property.

These assumptions of order are of particular interest historically in view of the fact that EUCLID makes abso-

[1] This theorem, which was given in the first edition of Hilbert's work as an assumption, was proved by E. H MOORE to be a consequence of the assumptions previously given. (*Transactions of the American Mathematical Society*, vol. III, 1902.)

lutely no mention of them, and that, as a result of this omission on the part of Euclid, it is possible, using his axioms and postulates alone, to derive many of the well-known paradoxes of geometry. We will give an example of one.

A Paradox. — *To "prove" that every triangle is isosceles.* Given any triangle ABC. We draw first the bisector of the angle A, and erect to BC a perpendicular at its middle point D. If these two lines were parallel, the bisector of the angle would be perpendicular to the side BC, and the triangle, by a well-known theorem, would be isosceles. If these two lines

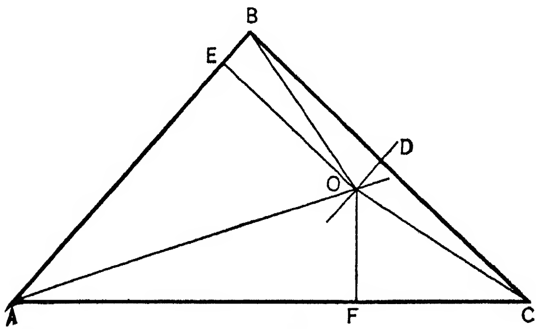

FIG. 15

are not parallel, they intersect in a point O, and we can distinguish two possible cases. The point O will lie either within the triangle ABC or without it. In either case, we draw straight lines OE and OF perpendicular to AB and AC respectively and join O to the points B and C. In the first case (Fig. 15), the two right triangles AOF and AOE are congruent, since they have the side AO in common and

their two angles at A are equal by construction. We have, therefore,

$$AF = AE, \text{ also } OF = OE.$$

Similarly, the two right triangles OCD and OBD are congruent, since they have the side OD in common and $DC = DB$ by construction. We have, therefore,

$$OC = OB.$$

From the latter equality, in connection with the equality $OF = OE$, already noted, we conclude that the two right

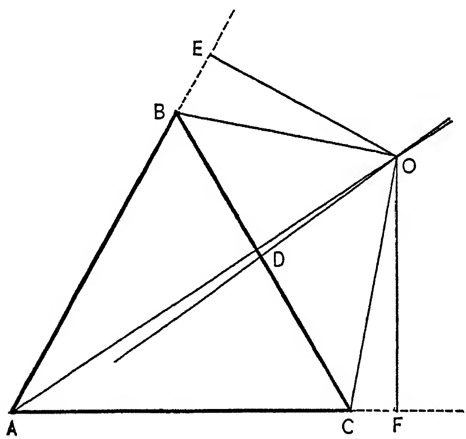

FIG. 16

triangles OCF and OBE are congruent; that, therefore, $EB = FC$. This proves, in connection with the equality $AE = AF$, by addition, the equality of the two sides AC and AB (Q. E. D.). In the second case (Fig. 16), if O is outside the triangle ABC, we may derive in precisely the

same way the congruence of three pairs of triangles, viz., $AOF \cong AOE$, $OCD \cong OBD$, and hence $OCF \cong OBE$, and find as before $AF = AE$ and $FC = BE$. By subtraction, we again obtain the result $AB = AC$, which proves the triangle is isosceles. The only thing that is wrong in this demonstration is the figures. According to Euclid's axioms and postulates alone, the demonstration is absolutely valid. As a matter of fact, the point O will always lie outside the triangle ABC, and, of the two points F and E, one will always lie within, the other will always lie without, the corresponding sides AC and AB of the triangle We then have, for example (Fig. 17),

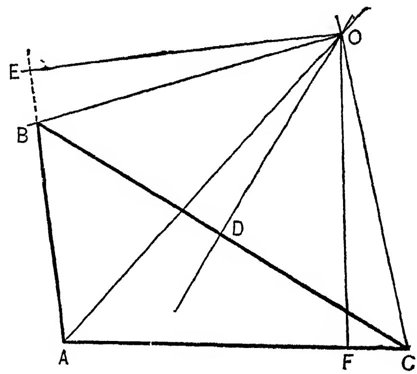

FIG. 17

$$AB = AE - BE, \text{ and } AC = AF + FC = AE + BE,$$

so that we cannot conclude that AB and AC are equal.

LECTURE XIV

HILBERT'S ASSUMPTIONS (*Continued*)

Assumptions of Congruence.—The third group of assumptions in Hilbert's set involve a new undefined relation between linear intervals and angles (the latter term will presently be defined), which is expressed by the word *congruent*. This relation is characterized by the following assumptions:

III. 1. *If A and B are two points of a straight line a, and A' is a point on the same or another straight line a', then there exists on a' in a given direction from A', one and only one point B' such that the interval AB is congruent with the interval $A'B'$. In symbols, $AB \cong A'B'$. Every interval is congruent with itself; that is, $AB \cong AB$, and $AB \cong BA$.*

III. 2. *If an interval AB is congruent with an interval $A'B'$, and also congruent with an interval $A''B''$, then $A'B'$ is congruent with $A''B''$. In symbols, if $AB \cong A'B'$, and $AB \cong A''B''$, then $A'B' \cong A''B''$.*

III. 3. *Given two intervals AB and BC on the straight line a, with no common points (other than B), and also two intervals $A'B'$ and $B'C'$ on the same or another straight line a', with no common points (other than B'): if $AB \cong A'B'$, and $BC \cong B'C'$, then also $AC \cong A'C'$.*

Definition of Angle.—These three assumptions are sufficient to characterize completely the congruence of linear intervals. Before proceeding further we must define angle. A point A of a straight line divides the straight line into two parts; each of these parts is called a *half-line* issuing from A. Given in a plane α two half-lines h and k issuing from a point O, and belonging to different straight lines, such a pair of half-lines h and k is called an *angle*, and is denoted by $\measuredangle\,(h, k)$ or $\measuredangle\,(k, h)$.

From assumptions II. 1 to II. 4 it can be readily proved that the half-lines h and k, together with the point O, divide the remaining points of the plane α into two regions characterized by the following properties. If A is any point of one of these regions, and B any point of the other, then any broken line joining A and B either passes through O or has at least one point in common with h or k; while, if A and A' are any two points of the same region, there always exists a broken line joining A and A' which neither passes through O nor has a point in common with the half-lines h and k. One of these regions is distinguished from the other by the property that any two points of this particular region determine an interval which is contained entirely in this region. This particular region is called the *interior* of the angle $\measuredangle\,(h, k)$, to distinguish it from the other region, which is called the *exterior* of the angle. The half-lines h, k are called the *sides* of the angle, and the point O is called its *vertex*. Assumptions III. 4 and III. 5 are entirely similar to assumptions III. 1 and III. 2, except that the word "angle" replaces the word "interval." The last assumption of this set is:

III. 6. *Given two triangles ABC and $A'B'C'$; if we have AB congruent with $A'B'$, AC congruent with $A'C'$, and angle BAC congruent with angle $B'A'C'$, then we also have angle ABC congruent with angle $A'B'C'$, and angle ACB congruent with angle $A'C'B'$.* From this set of assumptions follow, in particular, the well-known congruence theorems for triangles. It is also possible, with the assumptions thus far made, to define what is meant by a right angle, and prove as a theorem that all right angles are congruent, a proposition which, as we have seen, Euclid took as an axiom.

The Parallel Axiom. — The fourth set of assumptions consists of a single one, the so-called *parallel axiom.* It is as follows:

IV. *Let a be any straight line, and A any point not of a; then there exists in the plane α determined by a and A one and only one straight line b which contains A and does not meet a. This line is called the parallel to a through A.* As soon as this assumption is added to the preceding ones, many well-known theorems may be derived, among which we mention the following: If two parallels are cut by a third straight line, the alternate exterior and the alternate interior angles are equal, the corresponding angles are equal, etc. Conversely, the congruence of these angles implies that the straight lines are parallel; the sum of the angles of a triangle is equal to two right angles; etc.

The fifth and last group of assumptions is as follows:

Assumptions of Continuity. — V. 1. *The assumption of measurement (or the Archimedean assumption). Let A_1 be*

any point of a straight line lying between two given points A and B. Let A_2, A_3, A_4. A_5, … be a sequence of points, such that A_1 is between A and A_2, A_2 between A_1 and A_3, A_3 between A_2 and A_4, etc , and such that the intervals AA_1, A_1A_2, A_2B_3, A_3A_4, etc., are all congruent with each other; then there exists in this sequence of points a point A_n such that B is between A and A_n.

V. 2. *Assumption of completeness. The elements (points, straight lines, planes) form a class of objects, which under the set of all previously made assumptions, is not capable of further extension;* that is, it is not possible to add to the class of points, lines, and planes any other elements such that the new class satisfies all the assumptions I to IV and V. 1.

From these two assumptions, it is possible to prove that the points of a line form a linear continuum in the sense in which that term has previously been defined. Assumption V. 1 has become known as the axiom of Archimedes, in spite of the fact that it was known to Euclid long before Archimedes' time. It is of particular interest, since on it depends the whole theory of measurement, in particular Euclid's theory of proportion. In fact, Euclid states the axiom in the following form: *Two magnitudes are said to have a ratio, if they are such that a multiple of either may exceed the other.* It plays an important rôle in all modern investigations concerning the foundations of analysis and geometry.

A Non-Archimedean System. — We may get a clear notion of its bearing by considering a concrete example of a geo-

metric system for which it is not satisfied.[1] We know that in the ordinary method of measuring angles between straight lines this assumption is satisfied. If we consider angles between two curves, in the usual way, we define such an angle to be equal to the angle between the tangents to the curves at their point of intersection. Let us consider in particular the angle between two circular arcs. If the two circles intersect, without being tangent to each other, the angle, in the usual sense, has a completely defined measure greater than zero. If, however, the circles are tangent to each other, they form an angle of measure zero in the usual sense, since in that case the two tangents at their common point coincide. It is proposed now, however, to consider the angle between two intersecting circular arcs to consist of the figure itself in the immediate neighborhood of their point of intersection, and to show how, by a simple device, it is possible to arrange these angles in a linear order. For the sake of simplicity, let us confine ourselves to the consideration of angles of which one side is a fixed horizontal straight line a, and whose vertices are at a fixed point O of this line. The other side of the angle is then to be a circular arc, or, in special cases, also a straight line passing through O. We will consider only those portions of these circular arcs (or straight lines) which lie above the given straight line. Given two such "angles," it is then natural to define that one to be the smaller whose circular side lies below the circular side of

[1] This section follows closely the discussion in KLEIN, *Elementarmathematik vom höheren Standpunkte aus*, vol. II, p. 423.

the other (Fig. 18). According to this definition, for example, the "angle" formed with the given straight line by a circle which is tangent to this straight line will always be less than that formed by a circle which is not tangent; and, of two circles which are tangent, the one with the greater radius will form the smaller "angle." According to this convention, the set of all "angles" formed with the

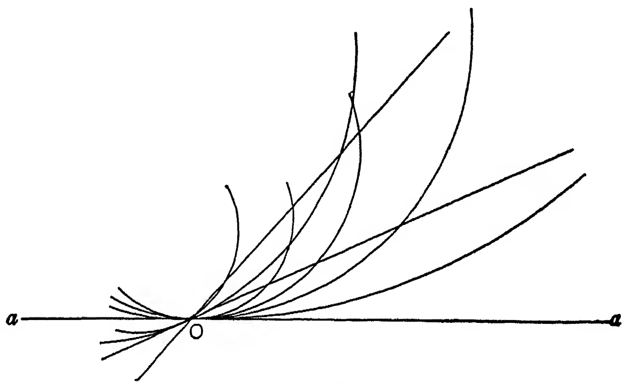

FIG. 18

given straight line at the point O by circular arcs and straight lines passing through O, is arranged in a linear order, which is readily seen to satisfy the three assumptions O_1, O_2, O_3, previously made to characterize such order. In order to exhibit the relation of this system of what we may call *lunar angles* with the axiom of Archimedes, it is necessary to define further what shall be meant by integral multiples of such an angle, or what shall be meant by multiplying one of these angles by a positive number n. Let us consider first the angle formed by a tangent circle of

radius r. It is natural to define as the n-fold multiple of this angle the lunar angle formed by the tangent circle whose radius is $\frac{r}{n}$ This is in accord with the previously made definition of order, in so far as the sequence of angles formed by tangent circles with the radii $r, \frac{r}{2}, \frac{r}{3}$, etc., become larger and larger; but it is seen at once that by this method the successive multiples of angles formed by tangent circles are all formed themselves by tangent circles, and therefore all such multiples, no matter how great n may be, are less, according to our definition, than the angle formed with the fixed line by any other straight line through O. It is therefore evident that the Archimedean axiom is not satisfied by this system of lunar angles. As for the addition of such angles, we may define the sum of the angles formed by two tangent circles as the angle formed by that tangent circle the reciprocal of whose radius is equal to the sum of the reciprocals of the radii of the two given circles. If then we have any circle through O which is not tangent to the given line, we may define the angle formed by this circle with the line as being the sum of the angles formed with this line by the tangent to the circle at O, in the ordinary sense, and the lunar angle between this point and the given circle. With these specifications, it is possible to apply the operations of addition and multiplication to any two such angles. Two such geometrical "magnitudes" then fail to satisfy Euclid's definition of magnitudes which have a ratio. We took occasion to call attention, in the last lecture, to the serious logical defect in Euclid's Ele-

ments of Geometry, which arose from his neglect to state the necessary assumptions regarding order. It is only just, therefore, that we call particular attention to his theory of proportion, from which modern criticism has detracted little or nothing.

Additional Remarks on Hilbert's Assumptions. — This, in brief outline, is the system of assumptions from which Hilbert succeeds in deriving the theorems of euclidean metric geometry. For further details regarding the actual derivation of the fundamental theorems, reference must be made to Hilbert's book, which we are all the more ready to do, since an English translation of this important work is available. It will be noted that the number of undefined terms (point, straight line, plane, between, congruent) and the number of assumptions, of which there are twenty-one, is rather large. It is large, at any rate, compared with certain other sets of assumptions which have been proposed, notably that given by Professor VEBLEN,[1] which contains only two undefined terms, "point" and "between," and contains only twelve assumptions. Hilbert's set, however, has the advantage, which is to be expected when the number of assumptions is comparatively large, of making the derivation of the fundamental theorems easier and more elementary. He proves that his assumptions are consistent and

[1] "A System of Axioms for Geometry," *Transactions of the American Mathematical Society*, vol. 5 (1904), p. 343. Reference should also be made to a more recent article by Professor VEBLEN, "The Foundations of Geometry," *Mathematical Monographs*, edited by J. W. A. Young, Longmans, Green and Co., 1911.

categorical, and also that they are independent in sets, that is, each of the sets I to V is independent of all the other sets, and all of the assumptions within a given set are independent among themselves; but he does not show that the whole set of twenty-one assumptions are mutually independent. Indeed, that such is not the case follows at once from the fact that the last assumption V. 2 is self-contradictory unless assumption V. 1 is also present.

Congruence vs. Motion. — Hilbert's set of assumptions is of interest further in that it chooses as one of the fundamental undefined notions the concept of congruence. There has been much discussion in the past regarding the question as to whether Euclid regarded the idea of congruence or the idea of motion as fundamental. The early part of his treatment seems to indicate that he desires expressly to avoid the notion of movement. This seems to follow particularly from his ingenious constructions for congruent segments; but very soon thereafter he draws conclusions which appear to be justified only on the hypothesis of a rigid displacement of the figures in question. The concept of rigid displacement is one of such fundamental importance that it seems desirable to devote the next lecture to the consideration of a set of assumptions in which this notion is taken as one of the undefined concepts.

LECTURE XV

PIERI'S ASSUMPTIONS

Undefined Terms : Point and Motion.—An Italian, MARIO PIERI,[1] has recently given a set of assumptions characterizing the ordinary metric geometry in which the notion of point and the notion of a rigid displacement or motion are the only undefined terms. The points are thought of as elements of a class S. Motion is considered merely as a one-to-one reciprocal correspondence between points, that is to say, as in the popular conception of the word "motion," the result of the motion alone is considered. The points and the motions are characterized by the following assumptions:

1. *The class* S *contains at least two distinct points.*

2. *Given any motion* μ *which establishes a correspondence between every point* P *and a point* P', *there exists another motion* μ^{-1}, *which makes every point* P' *correspond to* P. *The motion* μ^{-1} *is called the inverse of* μ.

[1] "Della Geometria elementare come sistema ipotetico-deduttivo ; monografia del punto e del mote" *Memorie della R. Academia delle Scienze di Torino (1899)*. We have followed the abstract of this paper given by L. COUTURAT, *Les Principes des Mathématiques*, Paris, 1905, p. 192.

3. *The resultant of two motions μ and ν performed successively is equivalent to a single motion.*

It follows from the last two assumptions that the class of all motions forms a group with respect to the operation of forming the resultant. It follows further from this assumption that the resultant of two motions μ and μ^{-1} is a motion which makes each point correspond to itself. This motion is called the *identical motion.* Any motion distinct from the identical motion we may call an *effective motion.*

4. *Given any two distinct points A and B, there exists an effective motion which leaves A and B fixed.* The effect of this assumption is to affirm the existence of the so-called rotations of a figure about two of its points.

5. *If there exists an effective motion which leaves fixed three points A, B, C, then every motion which leaves A and B fixed leaves C fixed.*

Straight Line Defined. — By means of this assumption it is possible to define the notion of straight line. Under the hypothesis of the last assumption the three points A, B, C are said to be collinear, and the *straight line* AB is then defined as the class of points consisting of A and B and all points collinear with A and B, that is to say, the class of points which remain fixed in any motion which leaves A and B fixed. It may be remarked that the straight line is hereby defined as a whole, and nothing in the definition implies any relation of order between its points.

It follows from the assumptions thus far made that a straight line is completely determined by any two of its points, that is to say. If C and D are two distinct points of

the line AB, then A and B are points of the straight line CD It is also of interest to note that the last assumption restricts the class of points to be three dimensional, a subject to which, as stated before, we shall return later. It is also possible to prove by means of these assumptions that if A, B, C are collinear, the points corresponding to them in any motion whatever are also collinear — in other words, that every motion transforms a straight line into a straight line.

Definition of Plane.—It is now possible to define a plane as follows: If A, B, C are three non-collinear points, the class of all points on the straight lines joining A to the points of BC, or B to the points of AC, or C to those of AB, is called the *plane ABC*. It is then possible to prove that all points of the straight lines AB, BC, CA belong to the plane, and that every motion transforms a plane into a plane; but it is not yet possible to prove that a plane is completely determined by three points. For this purpose is added the following assumption:

6. *If A, B, C are three non-collinear points and D is a point of the line BC distinct from B, the plane ABD is contained in the plane ABC* It is then possible to prove that the planes ABC and ABD coincide, and, more generally, that, if D, E, F are any three non-collinear points of the plane ABC, A, B, C are points of the plane DEF. It now follows also that a plane contains all the points of any straight line which contains two distinct points of the plane. It will be remarked that the latter proposition is often taken as the definition of a plane.

The Sphere—It is possible now to define a sphere as

follows: Given two distinct points A and B, the class of all points P, such that for every point P there exists a motion which leaves A fixed and transforms P into B, is called the *sphere* of center A and passing through B. It may be denoted by B_A It follows then that every motion transforms a sphere into a sphere, and every motion which leaves the center of a sphere fixed transforms the sphere into itself; also that, if two spheres with centers A and B have only one point C in common, the points A, B, C are collinear. In order to characterize the particular kind of motions which consist in turning a line about one of its points through an angle of 180°, the following assumptions are introduced:

7. *If A and B are two distinct points, there exists a motion which leaves A fixed and transforms B into another point of the straight line AB.* From this assumption may be proved the theorem that the sphere with center A passing through B has a second point in common with the straight line AB.

8. *If A and B are distinct points, and if two motions exist which leave A fixed and transform B into another point of the line AB, the latter point is the same for both motions.* This amounts to the statement that the sphere of center A and passing through B has only one other point in common with the straight line AB. This point, which we will call B', is thus uniquely determined by A and B, and is called the *opposite* of B with respect to A. It follows readily that the opposite of B' with respect to A is

9. *If A and B are two distinct points, there exists a motion which transforms A into B and which leaves fixed a point of*

the line AB. This motion also transforms B into A, and A and B are opposite to each other with respect to a fixed point of the line AB. This fixed point is uniquely determined, and is called the *mid-point* of A and B. It is the center of a sphere which passes through A and B, and is the only point of the line AB possessing this property. In order to characterize similar motions with respect to a plane, the following assumptions are necessary:

10. *If A, B, C are three non-collinear points, there exists a motion which leaves A and B fixed and which transforms C into another point of the plane ABC.*

11. *If A, B, C are three non-collinear points and D and E are points of the plane ABC common to the spheres C_A and C_B, and distinct from C, the two points D and E coincide;* in other words, there exists in the plane ABC only one point distinct from C which is at the same "distance" from C as the points A and B, provided A, B, C are non-collinear (if they are collinear, there is no such point). We may now define a *circle* as the class of all points common to a sphere and a plane passing through the center. The center of the sphere is also called the center of the circle. Assumptions 10 and 11 may then be stated in the following form: In a plane ABC, the circles C_A and C_B (a circle may be represented by the same notation as a sphere, provided the plane in which it lies is known) have one and only one point in common distinct from C. We may now define the perpendicularity of two straight lines as a relation between three points. The pair of points AC is said to be *perpendicular* to the pair AB, in symbols $AC \perp AB$, if there exists a

motion which leaves A and B fixed and transforms C into another point of the straight line CA. It is now possible to prove the theorem that from a point not on a line it is possible to draw one and only one perpendicular to that line In order to insure the existence of points exterior to a plane, we add the following assumption:

12. *If A, B, C are non-collinear points, there exists at least one point not in the plane ABC.*

13 *If A, B, C, D are four points not in the same plane, there exists a motion which leaves A and B fixed and which transforms D into a point of the plane ABC.* Assumptions 10 and 13 assert the possibility of a rotation It is then possible to prove the existence of translations, that is, motions which transform a plane into itself, which, at the same time, transform each of two parallel lines of this plane into itself Two points A and B are said to be *equidistant* from a point C, if C is the center of a sphere which passes through B and A. It is then possible to prove that the class of all points equidistant from two distinct points A and B in a plane containing A and B is the straight line perpendicular to AB at the mid-point of A and B, and hence is possible the definition of perpendicular between a straight line and a plane, etc.

Metric Relations. — These thirteen assumptions and the theorems which may be derived from them are concerned primarily with relations of position. It is now readily seen how it is possible from these relations of position to pass to relations of magnitude which form the chief object of study in metric geometry. A point is said to be *interior* to a

sphere if it is the mid-point of two distinct points of the sphere; it is said to be *exterior* in the opposite case, provided it is not a point of the sphere. In a plane a point is said to be interior or exterior to a circle according as it is interior or exterior to the sphere which has the same center as the circle and which passes through the circle. The sphere passing through two points A and B and whose center is the mid-point of AB, we will call the *polar sphere* of A and B. The point X is said to be *between* A and B, if it is a point of the line AB and is interior to the polar sphere of A and B. Finally, the class of all points situated between two distinct points A and B and including these points is called the *interval* AB (A and B are the *extremities* of the interval). It is thus seen how, by these purely logical definitions, it is possible to define the notions of "betweenness" and the notion of interval. We must now add the following three assumptions:

14. *If A, B, C, D are four distinct collinear points, the point D is not a point of one and only one of the intervals AB, AC, BC;* that is to say, the point D is either a point of none of these intervals or it is a point of at least two of them.

15. *If A, B, C are three collinear points, and C is between A and B, no point can be between A and C and between B and C at the same time.* In other words, the intervals AC and BC have no point in common except C. These two postulates serve to define linear order among the points of a straight line.

16. *If A, B, C are three non-collinear points, every straight*

line of the plane ABC which has a point in common with the interval AB must also have a point in common with the interval AC or the interval BC, provided the straight line does not pass through any of the points A, B, C. We find here again the assumption of Pasch, which was also taken in Hilbert's set (II. 4). It is now possible to define the extensions of an interval, the half-line, the half-plane, angle, etc, also the inequality of intervals, as follows: The interval AB is said to be *less than* an interval CD, if there exists a motion which transforms A into C and B into a point between C and D. Two intervals are said to be *congruent*, if there exists a motion which transforms one into the other. It is then possible to prove that of two intervals which are not congruent one is necessarily less than the other. The relation of order between angles is defined similarly. The *sum* of two intervals may now be readily defined in terms of motion. Thus it is possible to derive practically all the theorems of the first books of Euclid which are independent of the parallel postulate. For example, the base angles of an isosceles triangle are equal; if the angles of a triangle are unequal, the greater angle is opposite the greater side, and conversely; also, each side of a triangle is less than the sum of the other two. However, it is not possible to prove the following fundamental theorem: There exists a triangle whose sides are respectively equal to three given intervals of which each is smaller than the sum of the other two; or, what amounts to the same thing, two circles in the same plane intersect, if the distance between their centers is less than the sum of their

radii. In fact, this theorem presupposes a last assumption, the assumption of continuity. It is sufficient, moreover, to assume merely the continuity of linear intervals, as follows:

17. *If* C *is any class of points contained in the interval* AB, *there exists in this interval a point* X *such that no point of* C *is between* X *and* B, *and such that for every point* Y *between* A *and* X, *there is a point of* C *between* Y *and* X *or coincident with* X. From this assumption can be derived the assumption of Archimedes, which, as we have seen, forms the foundation for the measurement of magnitudes. Now, and now only, is it possible to affirm that linear intervals are measurable magnitudes.

Pedagogical Considerations. — We must content ourselves with this all too brief discussion of sets of assumptions characterizing metric geometry. We have chosen the two particular sets which have been outlined, because the one rests primarily on the fundamental notion of congruence, while the other is built up from the fundamental notion of motion. The current textbooks appear to combine these two notions in their treatment of geometry. In view of the fact that the notion of congruence, as such, appears to be much more abstract than that of motion, it would seem that the latter should be made the fundamental one also in our elementary texts, and congruence defined in terms of it, as indeed is usually the case. (Two figures are said to be congruent if one can be superposed upon the other so as to coincide throughout.) This does not mean, of course, that the set of assumptions given by Pieri should find a place in

our elementary textbooks; in particular, it does not mean that we should define a straight line joining two points as the class of points which remain fixed under any motion which leaves the two points fixed. We need only recall what was said in a previous lecture. The teacher's problem is, especially at the beginning, far more psychological than logical. It is obvious that a subject must be presented to a boy of fourteen years in a different way from that employed in presenting it to a mature mind. In particular, it is necessary, in the beginning, to make continued and insistent appeal to concrete geometric intuition. From this point of view, the notions of point, straight line, and plane may be assumed to be sufficiently clear in the pupil's mind without any formal definition. Moreover, it hardly seems necessary to say that many assumptions which are essential in a purely formal logical development of the science may and should be tacitly assumed in a first course. This is merely another form of the assertion that the power of abstraction and the amount of formal reasoning expected of a pupil at a given time must be adapted to his capacity to form such abstractions and formal deductions at that time. His capacity in this direction will slowly but surely increase, if it is allowed to develop naturally; it will be greatly impaired, if not altogether destroyed, by any attempt to force its growth.

The best way to take account of this psychological element would seem to be the removal of all formal considerations from the beginning of the course in geometry: as much as a half of the first course might profitably be

devoted to an informal treatment of geometry, in which the pupil is made familiar with the more important figures and constructions, and in which he is encouraged to think about the problems which present themselves in his own way. During this part of the course the pupil could be led to see the advantages of the more formal methods that follow. Unfortunately none of our present textbooks provide for such an informal introduction.

LECTURE XVI

THE DIMENSIONS OF A CLASS

The Three Dimensions of Space — The familiar proposition that space is three-dimensional, if analyzed, is seen to mean about the following: If a point moves along a straight line, it generates a space of one dimension; if this straight line in turn moves in a direction differing from its own, it generates a space of two dimensions, or a surface; if this surface in turn moves in a proper direction, it generates a space of three dimensions; and this exhausts all the points of space as we know it. The fact that three motions are necessary in this process is what is usually expressed by the statement that space has three dimensions. It follows at once, from this description, that any point in space may be represented by three numbers (coordinates), the first of which fixes the position of the point in question on a line, the second of which determines the position of this line on a surface, and the third of which determines the position of this surface in space. This is the fundamental idea of analytic geometry, to which we shall return later.

Dimensionality and Cardinal Number. — If we attempt to describe it in purely logical terms, that is, without making

use of the more or less vague notions of motion, direction, etc., employed in the above description, we are first led to the question as to whether the number of dimensions of a class is a property which can be thought of as pertaining to the class as such, relating in particular to the cardinal number belonging to such a class. Nothing is more natural, from the purely intuitional point of view, than to suppose that, if a certain infinite number belongs to the class of points on a line, an infinitely greater infinite number belongs to the class of points in a plane. That such is not the case, however, was first recognized by CANTOR, and has since been further established by the investigations of PEANO, HILBERT,[1] E. H. MOORE,[2] and others. The fact is that the cardinal number of the points on a line is precisely the same as the cardinal number of the points in a plane or the points in space; in other words, that it is possible to establish a one-to-one reciprocal correspondence between the points of a line and the points of a plane (or of space). A general idea of how such a correspondence can be established may perhaps be obtained from the following description, which is a general outline, omitting the necessary details, of the proof due to Hilbert.

Correspondence between the Points of a Line Segment and of a Square. — We consider a square whose sides are of given length, and a straight line of given length, and will show

[1] HILBERT, "Über die stetige Abbildung einer Linie auf ein Flächenstuck," *Mathematische Annalen*, vol. 38, 1891, pp. 459-460

[2] E. H. MOORE, "On Certain Crinkly Curves," *Transactions of the American Mathematical Society*, vol 1, 1900, p. 72.

how to establish a one-to-one reciprocal correspondence between the points of the square and the points of the line. For this purpose, divide each side of the square into two equal parts, and by joining the points of division, divide the square into four equal squares. Corresponding hereto divide the given straight line into four equal parts. Let the squares be numbered in order, 1, 2, 3, 4, in such a way that squares with two successive numbers have a common side. This may be illustrated by drawing a broken line, which starts at the middle of one of the squares and passes successively through the middles of adjacent squares (Fig. 19). Corresponding to the numbering of these four squares, let us number the four intervals on the given line 1, 2, 3, 4; and let each one of the four squares 1, 2, 3, 4 cor-

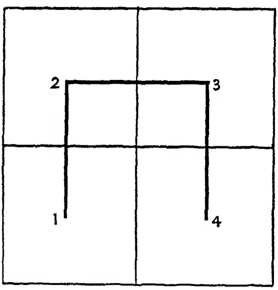

FIG. 19

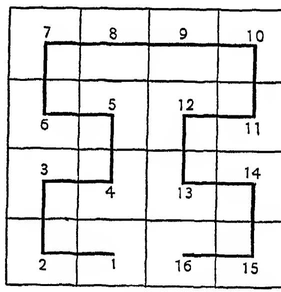

FIG. 20

respond to the interval with the same number on the line. Then let us treat each of the four smaller squares and each of the intervals on the line in precisely the same way; that is, divide each of the four squares into four new squares, each of the four intervals on the line into four new

intervals. We obtain in this way $16 = 4^2$ squares and the same number of intervals, which may again be numbered 1, 2, 3, · · ·, 16 (Fig. 20), and put into correspondence in such a way that (1) to two distinct squares correspond two distinct intervals and (2) to four squares comprising one of the larger squares there correspond the four intervals of the interval which corresponds to the larger square. It is clear that this process can be continued indefinitely, with the result that after the nth stage has been reached the original square has been divided into 4^n small squares and the given line into the same number of intervals, such that the 4^n squares into which the given square is divided correspond in a one-to-one way to the 4^n intervals of the given line, so that the properties (1) and (2) are preserved, Then to any infinite sequence S_n of squares ($n = 1, 2, 3, \cdots$), in which every square S_n includes the succeeding square S_{n+1}, there corresponds an infinity of sequences of intervals I_n, in which every interval I_n includes the succeeding interval I_{n+1}. It may then be shown by means of the geometric axioms of continuity in the line and plane that any such sequence of squares determines a unique point in the square, and the corresponding sequence of intervals I_n determines a unique point of the line. Moreover, every point of the square and every point of the line may be determined by such an infinite sequence of squares and intervals. This being granted without proof, it is seen that there is indeed established a one-to-one reciprocal correspondence between the points of the given square and the points of the given line; that is, the cardinal

number of the points in the square is equal to the cardinal number of the points on the line.

Dimensionality a Property of Order. — It follows from this theorem that the dimensionality of a class is not an inherent property of the class as such. We shall see at once that it is a property of order in a class. Indeed, since, as we have seen on several occasions, the elements of a class may be ordered in different ways, it will follow that the dimensionality of one and the same class may be different, depending on the way in which it is ordered To come immediately to the point, a class which satisfies the assumption for linear order is said to be one-dimensional. A two-dimensional class is then a linearly ordered class of linearly ordered classes. This may be called a doubly ordered class. A linearly ordered class may now conveniently be called simply ordered. A three-dimensional class is then a simply ordered class of doubly ordered classes. The important fact to be noted is that when we think of a class as two-dimensional, for example, we are thinking, strictly speaking, not of a class, as such, but of a *class of relations*. To illustrate: The points of a plane, when thought of as two-dimensional, in the ordinary way, is two-dimensional by virtue of the fact that it consists of a simply ordered class of lines, each line being a simply ordered class of points. Space, when considered as three-dimensional in the ordinary way, is three-dimensional because it is a simply ordered class of planes, each of which in turn is a simply ordered class of lines, which are composed of simply ordered classes of points.

A Four-Dimensional Class. — With this conception of dimensionality, classes of any number of dimensions are readily seen to exist. The class of all spheres in space, for example, is a four-dimensional class. To see this, we think of all spheres with a given center forming a simply ordered class C_1 (ordered according to the length of their radii), then think of the simply ordered class C_2 of such classes C_1 obtained by allowing the center to move on a straight line, then of the simply ordered class C_3 of such classes C_2 obtained by letting this line of centers describe a plane, and finally by thinking of the simply ordered class C_4 of such classes C_3 obtained by allowing this plane of centers to describe the whole of space.

Coordinates and Analytic Geometry. — It follows at once from this definition of dimensionality that the elements of a class of n dimensions can always be represented, labeled as it were, by sets of n numbers called coordinates, each number of such a set determining an element of the n linearly ordered classes of which the whole class is composed An illustration of this fact has already been referred to in the case of the representation of the points of space by means of three numbers (coordinates). In the last example given, that of the four-dimensional class of all spheres in space, it is clear that any such sphere is uniquely determined if the three numbers locating its center and the number measuring its radius are given.

These considerations lead naturally to the introduction of the ideas lying at the basis of so-called analytic geometry, or conversely, of geometric interpretations in algebra. Let

us confine ourselves for the sake of brevity to the two-dimensional case. The method of representing the points of a plane by pairs of numbers x and y is well known. In the usual cartesian system, the two numbers x and y are used to represent the distances and directions of the point in question from two fixed mutually perpendicular lines of the plane taken as lines (axes) of reference. Any relations between the points in a plane then gives rise to relations between pairs of numbers representing them, and, conversely, to any relations between such pairs of numbers there will correspond geometric relations between the corresponding points. This is not the place to go into any further details regarding the methods and results of analytic geometry. It may, however, be desirable at this point to emphasize the importance of the judicious use of such graphic representation in the study of algebra. It is one of the best means of furnishing the pupil with a concrete representation of the symbolism which he is studying and to which reference has previously been made.

It may not be out of place at this point to say a word regarding the conception of a space of four or more dimensions. This subject, which appears to have aroused much interest during recent years in the popular mind, will be considered briefly in the next lecture.

LECTURE XVII

SPACES OF FOUR OR MORE DIMENSIONS

Four-Dimensional Space. — We have already seen that there is nothing at all mysterious in the conception of a class of four or more dimensions; moreover, that the elements of such a class may be geometrical. As an example, we may recall the four-dimensional class of all spheres in space, which was referred to in the last lecture. That such classes form a legitimate and valuable object of study in geometry is not open to question, and if we agree to apply the word "space" to such a class, we at once see the possibility of geometric spaces of any number of dimensions. These spaces are, of course, not spaces of points, however. There still remains, then, a question as to whether the conception of a point space of four dimensions is justifiable. Our concrete intuition of space makes it three-dimensional. We find it impossible to conceive concretely of any further points. From the purely abstract point of view developed in these lectures, however, no such concrete representation is required (except for the purpose of consistency proofs). Nothing can prevent us from assuming the existence of further points, provided such an assumption does not con-

tradict any of our previous geometrical assumptions. We have seen in both sets of assumptions for geometry which we have discussed that they contained an assumption which implied that the class of points considered was three-dimensional. If these assumptions had been replaced by others not having this implication, the class of points characterized by these assumptions might well have been more than three-dimensional, and the only justification necessary for studying the logical implications of such a set of assumptions is that of logical consistency. That logically consistent sets of assumptions characterizing a four-dimensional class of points are indeed possible is readily shown, and will indeed appear presently in this lecture. The desirability of studying such an abstract space of four or more dimensions is of course simply dependent upon the interest and serviceableness which attaches to the results obtained. It may be stated, without fear of contradiction, that the study of such spaces has been of the greatest practical value both in pure mathematics and in the applications of mathematics to the physical sciences.

Spaces of n Dimensions in Applied Mathematics. — Indeed, it is a fact, which may at first appear strange, that the study of geometries of more than three dimensions was made almost necessary by certain problems in applied mathematics. How this came about may be suggested by the following considerations. The analytic formulation of the problem of describing the motion of a body in a straight line involves a relation between a variable x determining the position of the body on the line at a given time

and the time t. It is often desirable to represent this relation graphically, by interpreting any pair of simultaneous values of t and x as a point in a plane, according to the methods of analytic geometry recently referred to. This simplest sort of a mechanical problem requires for its geometric interpretation the points of a plane. If we consider the motion of a body in a plane, we at once have a relation between three variables, the coordinates x and y of the point and the time t; and any system of simultaneous values of t, x, y may be interpreted as a point in space, an interpretation which is often of the greatest service. If now we consider the fact that the motion of even a single body in space involves four variables, the time t and the corresponding values of the coordinates x, y, z of the body, it is at once seen that our three-dimensional space of points will not admit of an analogous geometric interpretation. The problem demands that we consider, as in the other case, simultaneous sets of values of four variables t, x, y, z. These four variables are subject to certain equations, and the analytic discussion of the mechanical problem in question requires the discussion of such a set of equations involving four variables. In general our problems of mechanics are concerned, however, with more than one moving body. If we consider n such bodies, we find it necessary to consider, in addition to the time t, the $3n$ coordinates determining the position of the set of n bodies in space, and must then consider equations involving these $3n+1$ variables and sets of simultaneous values of such $3n+1$ variables. We have just noted that the concrete geometric

interpretation of such a simultaneous set of values by points in space is not possible. But even if we give up entirely the possibility of such a geometric interpretation, the language necessary to discuss the analytic problem in question soon becomes extremely cumbersome, unless some convenient term is adopted to denote sets of simultaneous values of the variables in question. What is more natural, then, than to call such a set of simultaneous values a *point?* Such a terminology does not necessarily imply any geometric notion at all, but may be employed merely for the sake of its convenience. According to our previous definition the class of all such "points" is more than three-dimensional. However, the adoption of such a terminology at once suggests analogies with ordinary geometry, analogies which often prove to be of the greatest service in disentangling the complex relations defined by the equations characterizing the problem.

Analytic Space of n Dimensions.—It was therefore inevitable that the study of such an "analytic space" of any number of dimensions should be made with the same methods and with a terminology similar to that belonging to the ordinary analytic geometries of two and three dimensions. Such a study having proven itself of value in physical applications, it was then an easy step to study it for its own sake. It soon appeared that the consideration of the properties of an abstract point space of four or more dimensions threw more light upon the logical interrelations of figures in three-dimensional space. It seems desirable, therefore, to consider the four-dimensional space of points

from the abstract point of view a little more fully, with a view of justifying the last assertion.

Abstract Point Space of Four Dimensions. — If we postulate the existence of a four-dimensional space of points, it will, of course, be desirable to choose the postulates in such a way that our ordinary geometry of three dimensions will form a part of this more extensive four-dimensional geometry, just as the ordinary plane geometry forms a part of the familiar geometry of three dimensions. Just as there exists in three-dimensional space an infinite number of planes, we should think of an infinite number of three-dimensional spaces — *three-spaces*, we will call them — in the more extensive space of four dimensions. To set up a set of assumptions for such a space, it would then be natural to add to the undefined terms "point," "line," and "plane," an additional undefined term "three-space." The assumption I. 7 of Hilbert, which we noted previously as being the one which implied the three-dimensionality of space, should then be modified to read: If two planes *are in the same three-space* and have a point A in common, they have at least one other point B in common. Other assumptions would have to be added, characterizing the theorems of alignment within a three-space, which in the previous set could be proved as theorems, such as, for example: If a plane has three non-collinear points A, B, C in common with a three-space, all of its points are points of the three-space. A three-space is determined by any four of its points which are not in the same plane. In order to insure the existence of our four-dimensional space we should have to add an assumption analogous to I. 8:

There exist at least five points not lying in the same three-space. Corresponding to I. 7, already referred to, we should have the assumption: If two three-spaces have a point A in common, they have at least two other distinct points B and C in common, which are not collinear with A. This assumption will imply that the space considered is not more than four-dimensional. These examples of the new assumptions must suffice to give a general idea of the nature of the new assumptions necessary to characterize a space of four dimensions. That such a set of assumptions may be chosen so as to be logically consistent is then readily shown by constructing the analytic space of the kind described above, in which each point consists of a set of four numbers, and in which the straight lines, planes, and three-spaces are defined by means of certain equations. We have already referred to the fact that in such a four-dimensional space the geometry within any three-space will be precisely the same as the three-dimensional geometry with which we are familiar. The new element which enters is merely the fact that any such three-dimensional geometry is merely one of an infinite number of such geometries, all equivalent to each other, just as the geometry in any plane is merely one of many, all equivalent to one another, within a three-dimensional space. However, just as the geometry in a plane bears certain relations to the geometry of the three-dimensional space of which it forms a part, so the geometry in a three-space will bear certain relations to the geometry of the four-dimensional space within which it lies; and we are now in a position to indicate how this new conception

may throw additional light on theorems of ordinary three-dimensional geometry.

Value of a Four-Dimensional Space in Studying Geometry. — Consider for example two triangles ABC and $A'B'C'$ (Fig. 21) in a plane, one of which may be obtained from the

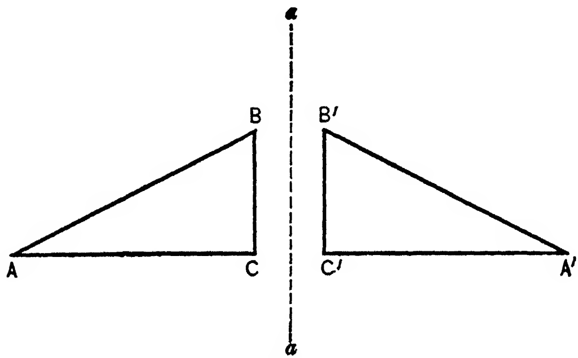

FIG. 21

other by rotating the plane through an angle of 180° about an axis a in the plane. We are in the habit of saying that these two triangles are congruent. If, however, we were confined to motion in a plane, it would be impossible to superpose one of these triangles upon the other, or to move one of the triangles in such a way as to bring it into coincidence with the other. It is only by going outside the plane and turning one of the triangles over that it is possible to superpose them. Let us now consider an analogous situation in space. Suppose we have two tetrahedra $ABCD$ and $A'B'C'D'$ (Fig. 22), which are symmetric, that is, "mirror images" of each other. We refuse to call them congruent

for the simple reason that no motion in three-dimensional space is possible which will bring one into coincidence with the other. If, however, we consider the three-space in which they lie to be part of a four-dimensional space of the kind we have described above, then it can be proved

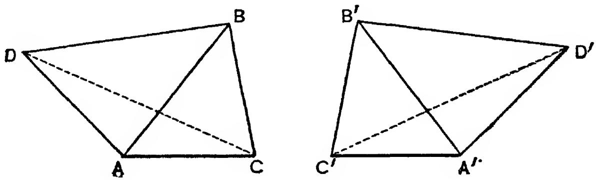

FIG 22

from the assumptions characterizing such a space that there exists a "motion" which will carry one of these tetrahedra into the other. From this point of view, there is no more reason to regard two such tetrahedra as non-congruent than there is to regard the two triangles mentioned above as non-congruent. This is only one of many examples which might be given to show how the conception of an abstract four-dimensional space can serve to exhibit new and interesting relations within the ordinary three-dimensional geometry.

LECTURE XVIII

ALGEBRA AND GEOMETRY

Definition of Certain Branches of Mathematics. — It was stated in the first lecture that to give a satisfactory definition of mathematics as a whole is difficult, if not impossible. We are, however, in a position now to see that it is a simple matter to define certain branches of mathematics. Any such branch, as we have seen, is completely characterized by means of a set of assumptions. Ordinary real algebra may be defined simply as the class of all formal logical implications of the set of assumptions given in Lecture XI characterizing the real number system. Similarly, ordinary euclidean metric geometry may be defined as the class of all formal logical implications of a set of assumptions such as Hilbert's.

Algebra and Geometry Abstractly Equivalent. — It is a very interesting and significant fact, however, that from this formal abstract point of view the content of ordinary real algebra and ordinary metric euclidean geometry coincide absolutely, so that the distinction between these so-called "branches" is entirely wiped out. That this is so is readily seen. We saw in the last lecture how it is possible

to represent the points of space by sets of three coördinates. It is therefore possible, starting with the set of assumptions characterizing the algebra of real numbers, to define a system of things which is abstractly equivalent to metric euclidean geometry. We merely *define* a "point" to be such a set of three real numbers, x, y, z, and consider the class of all points (that is, three-dimensional space) as consisting of all such sets. The plane is then *defined* as the set of all these "points" which satisfy an equation of the first degree.

$$ax + by + cz + d = 0,$$

in the coordinates x, y, z; and the straight line is simply *defined* as the class of all "points" common to two such "planes," provided they have "points" in common. The assumptions of alignment may then all be shown to be satisfied by the "points," "straight lines," and "planes" of this system. The further notions of "betweenness" and congruence are also readily defined, and the assumptions of parallelism and continuity shown to hold. We see then that, abstractly considered, *the whole content of metric euclidean geometry of three dimensions (and indeed of any number of dimensions) is contained among the implications of the set of assumptions defining the ordinary algebra of real numbers.* The converse process is also possible, moreover. Starting with a set of assumptions characterizing metric euclidean geometry of three dimensions, we may define the real numbers as the system of all intervals on a straight line, having a given point as one extremity, and define addition and multiplication of these intervals by means of

simple geometric constructions. It can then be proved that the operations thus defined satisfy the assumptions characterizing these operations in a number system, and the other assumptions necessary to characterize a system of real numbers may be seen to hold for this system of intervals. Consequently the whole content of real algebra considered abstractly is contained among the logical implications of a system of assumptions characterizing geometry. *The two branches, algebra and geometry, are then indeed seen to be abstractly identical in the sense that either includes the other.* Their difference consists merely in the arrangement of the propositions in a logical sequence. The recognition of this identity at once makes clear the many interrelations between algebra and geometry with which we are familiar. Algebraic methods are applicable to geometric problems; geometric methods are applicable to algebraic problems.

It seems desirable to illustrate this fact by an important example, especially since the ideas involved in the example chosen are all too little known among teachers of elementary mathematics. Construction problems form one of the most important and the most interesting objects of study in our courses in geometry. The pupil is shown how to inscribe in a circle regular polygons of three, four, five, and six sides; but when he, or his teacher, attempts to give a construction by means of a ruler and compass for the regular inscribed heptagon (that is, a figure of seven sides), he fails. Why? Simply because the construction of such a figure with the ruler and compass is impossible. The proof that this is so will furnish us an interesting example

of the interrelation between algebra and geometry, and, in particular, of the geometric interpretation of complex numbers described in Lecture XII.

Constructions with Ruler and Compass. — It is necessary, first of all, to determine with precision the limitations imposed on our constructions by the use of the straight-edge and compass alone. It is well known that with these instruments we may construct a segment equal to a given segment or a segment which bears any commensurable ratio to a given segment, or finally, that it is possible with a ruler and compass to construct a segment whose measure is equal to the square root of the measure of any given segment. The latter problem involves merely the construction of a mean proportional between the given segment of length a and a segment of length 1. The required segment, having the length x, is defined by the proportion $a:x::x:1$, which gives $x=\sqrt{a}$. The construction is one with which we are thoroughly familiar. By the combination of such constructions it is therefore possible, starting with any given segments of lengths a, b, c, etc., to construct any other segment whose length is represented by a number which is obtainable from the given numbers, a, b, c, etc., by any finite number of additions, subtractions, multiplications, divisions, and extractions of square roots. However, it may easily be shown that this includes all possible segments that can be constructed from a given set by means of a ruler and compass. This follows from the fact that a straight line and a circle are represented in analytic geometry by equations of the first and second degrees,

respectively, and that the determination of the intersections of circles with straight lines, or with other circles, leads analytically merely to the solution of quadratic equations, which involves no irrational operations other than the extraction of square roots. *If, then, we can show of any given geometrical problem that it involves the construction of a segment whose measure is obtained from the measure of the given segments of the problem by operations other than the rational and the extraction of square roots, we shall have proved that the problem in question cannot be solved with the ruler and compass.*

The Regular Pentagon. — After these preliminary remarks, let us return to the problem of inscribing in a circle a regular polygon of a given number of sides. If we take first the case of the regular pentagon, we may see why it is possible to construct it with the ruler and compass. Let the system of ordinary real and complex numbers be represented by the points of a plane as described in a previous lecture (XII). Let the circle in question have its center at the point O (Fig. 23) of this plane, and let its radius be unity. Suppose the point A_0 (representing the number 1) is one vertex of the pentagon to be inscribed; the other vertices A_1, A_2, A_3, A_4, will then be distributed on the circumference of the circle at equal intervals. Let the point A_1 be the representative of the complex number z. According to the interpretation of multiplication in this representation of complex numbers, it will be recalled that the product of any two complex numbers a and b, represented by points A and B respectively, is represented by the point C, whose

distance from O is measured by the product of the two numbers representing the distances of A and B from O, and which is such that the line joining C to O makes an angle with the real axis equal to the sum of the two angles

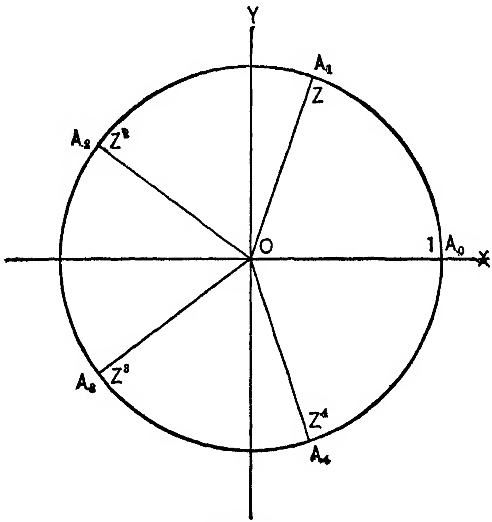

FIG. 23

made with this axis by the lines OA and OB. The product of any two numbers represented by points on the circle just described is therefore represented by another point on this circle, since in this case the distances from O are by hypothesis equal to unity. If we consider, in particular, the point z, we see that the number z^2 will be represented by the point A_2, since the angle XOA_2 is equal to twice the angle XOA_1. Similarly, the number z^3 is represented by the point A_3, the number z^4 by the point A_4, and finally

the number z^5 by the point A_0. This proves that the number z representing the point A_1 satisfies the equation

$$(1) \qquad z^5 - 1 = 0.$$

The left member of this equation is divisible by $z-1$, which yields the root 1, or the point A_0, in which we are not interested. The quotient of this division is the equation of the fourth degree

$$(2) \qquad z^4 + z^3 + z^2 + z + 1 = 0.$$

This equation is one of the type known as reciprocal, since, if any number α is a root of this equation, $\frac{1}{\alpha}$ is likewise a root. All the roots of this equation can be expressed in terms of the rational operations and the extraction of square roots. It is for this reason that the construction of a regular inscribed pentagon is possible with the ruler and compass. To prove the statement just made regarding the roots of this equation, we may suppose

$$(3) \qquad z + \frac{1}{z} = x,$$

from which follows, by squaring,

$$z^2 + \frac{1}{z^2} = x^2 - 2,$$

and if then we write equation (2) in the form

$$z^2 + z + 1 + \frac{1}{z} + \frac{1}{z^2} = 0,$$

and substitute from the relations just obtained, we find

$$x^2 + x - 1 = 0,$$

which is a quadratic equation for determining x, and if x is known, (3) gives a quadratic equation for the determination of z. No irrational operations other than the extraction of square roots are necessary to solve this equation.

The Regular Heptagon.—The treatment of the inscribed heptagon is entirely similar to this, although the result, from the point of view of geometric constructions, is radically different. If, as before, we take the circle in question to be the circle with unit radius and with its center at the

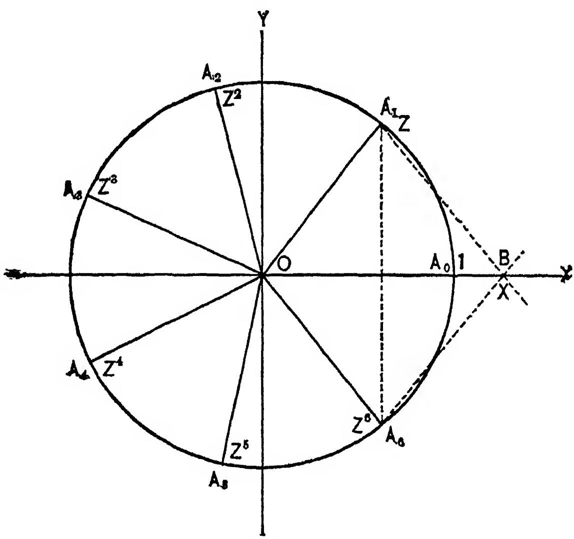

FIG. 24

point O (Fig. 24) of the complex plane, and choose the point $1 = A_0$ to be one vertex of the inscribed heptagon sought, the other vertices A_1, A_2, A_3, A_4, A_5, A_6, are seen to be repre-

sented by the complex numbers z, z^2, z^3, z^4, z^5, z^6, and we find that

$$(4) \qquad z^7 - 1 = 0.$$

The left-hand member of this equation is also divisible by $z-1$, the quotient in this case being

$$(5) \qquad z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0,$$

which is another reciprocal equation. In order to consider the solution of this equation of the sixth degree, we again put

$$z + \frac{1}{z} = x,$$

and find

$$z^2 + \frac{1}{z^2} = x^2 - 2, \qquad z^3 + \frac{1}{z^3} = x^3 - 3x,$$

so that if equation (4) is put into the form

$$z^3 + z^2 + z + 1 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} = 0,$$

x may be determined as the solution of the cubic equation,

$$(5) \qquad x^3 + x^2 - 2x - 1 = 0.$$

As before, if a solution of this equation is known or can be constructed, the calculation of z in terms of x will involve only the extraction of square roots. However, a radical difference between the present problem and the preceding one exists in the fact that equation (5) cannot be solved by means of rational operations and extractions of square roots alone. If it were possible to construct the point B representing the number x, the construction of the point A_1 representing the number z would follow very easily, as may be seen by consulting the figure. Indeed, the perpendicular

bisector of the segment Ox intersects the unit circle in the required point A_1. To prove this, we notice first that the point A_6, representing the number z^6, also represents the number $\frac{1}{z}$, for upon dividing both members of equation (4) by z we obtain

$$z^6 = \frac{1}{z}.$$

To form the sum of the numbers z, $\frac{1}{z}$, represented by the points A_1, A_6, respectively, we have only to complete the parallelogram of which OA_1 and OA_6 are adjacent sides and A_1A_6 is one diagonal, in accordance with the rule for the addition of complex numbers (described in Lecture XII). The extremity of the other diagonal is evidently the point B in question, representing the number

$$z + \frac{1}{z} = x.$$

The Trisection of an Angle.—A much simpler example of this sort of proof is furnished by the problem of trisecting an angle. Suppose the angle in question is placed with its vertex at the zero point O of the complex plane, and that one of its sides extends along the positive end of the x axis. Suppose the other side cuts the circle with unit radius and center at O in a point which is the representative of the complex number a. It then readily follows from considerations similar to those just noted that the side of an angle equal to one third of the given angle will meet the unit circle in a point x satisfying the equation

$$x^3 = a.$$

The determination of the value of x is, therefore, equivalent to the solution of this equation, which in turn involves the construction of a cube root, an operation which is in general impossible with the ruler and compass alone.

We observed in a previous lecture that the number π does not satisfy any algebraic equation with rational coefficients. It follows from this that the number π representing the ratio of the circumference of a circle to its diameter cannot be constructed with the ruler and compass, in other words, that the famous problem of the squaring of the circle by means of ruler and compass is as impossible as that of the construction of a regular heptagon inscribed in a circle, or the trisection of an angle.[1]

[1] For a more complete discussion of these problems, the reader is referred to KLEIN, *Famous Problems of Elementary Geometry*, translation by BEMAN and SMITH, Boston.

LECTURE XIX

VARIABLE. FUNCTION

Variable and Function. — Our discussion of the fundamental concepts of mathematics hitherto has been confined almost exclusively to those immediately concerned with the logical foundations. If, however, we glance at the content of elementary mathematics, we note that there are other derived concepts which have not been considered. Among these, perhaps the most important are the notions of variable, function, and limit. It is proposed to discuss them briefly, in the next two lectures.

The notion of functionality is one of such fundamental importance in our daily life that it should be introduced as early as possible into our school curriculum. The popular definition of a function, viz., that a quantity y is said to be a function of another quantity x, if y depends upon x, finds abundant illustration, not merely in the quantitative expression of the simplest physical laws, but in a more general sense in many of our daily activities.

The mathematical definition of function presupposes the notion of a variable. If this notion is analyzed, its essential characteristics will be found in the following definition: A *variable* is a symbol which represents any one of a class

of elements. This definition is very general; the elements of the class may or may not be numbers. The usual definition of a variable as being a number which varies is not only unnecessarily restrictive, but has the defect of introducing an extraneous and rather vague notion, that of varying or changing. It seems logically more desirable to consider a number as an element of a class. Such an element is not conceived to change, should not be thought of as changing. The fundamental idea at the basis of the notion of a variable, after all, is merely that we think, not of a particular element of a class, but rather of any one of a number of such elements. The definition of a variable which we have given clearly brings out this conception and entirely avoids the logically complex notion of an element changing in some way or other. It may be noted, in passing, that this definition of a variable is of distinctly recent date.[1] According to this definition a *constant* is a special case of a variable in which the class in question contains only a single element. Any particular element represented by a variable x may then be called a *value* of the variable.

We are now in a position to give a mathematical definition of a function. A variable y is said to be a *one-valued function* of a second variable x, provided there exists a correspondence between the class of the variable x and the class of the variable y such that to every value of x there corresponds a unique value of y.

[1] It seems to have appeared explicitly in a somewhat less general form for the first time in VEBLEN and LENNES, *Infinitesimal Analysis*, New York, 1907, p 44.

History of the Concept of a Function. — This conception of a function is a very general one, and was reached only after many years of gradual extension. Indeed, the history of the term *function* forms an interesting example of the tendency of mathematics to generalize its concepts. The word *function* appears to have been used first by DESCARTES, in 1637. He used the word simply to mean an integral power of a variable x such as x^2, x^3, etc., and indeed such expressions are functions in the modern sense defined above. This meaning seems to have been attached to the term until the time of LEIBNITZ (1646–1716). He used it apparently to denote any quantity connected with a curve, such as the coördinates of a point on the curve, the length of a tangent, etc. In this meaning of the term, the notion of correspondence or of the dependence of one variable upon the other is slightly obscure, but it is likely that Leibnitz thought of such quantities connected with the curve as being interrelated; that is, as being functions of one another in the sense above defined. JOHN BERNOULLI (1667–1748) defined a function to be any expression made up of one variable and any constants whatever. A little later, about the middle of the eighteenth century, EULER (1707–1783) called the functions defined by Bernoulli *analytic functions*, and made a distinction, which does not appear to be very clear, between algebraic and transcendental functions. The distinction was probably that between a function defined by an algebraic equation and one defined in some other way. This conception of function remained unchanged until the time when FOURIER (1768–1830), near the beginning of the

nineteenth century, began his investigations regarding the problems of vibrating strings, flow of heat, etc., which led him to the use of so-called trigonometric series. It was soon found that these series defined a more general type of correspondence than any before discussed; and, in the attempt to formulate a definition of function sufficiently general to include all such types, LEJEUNE DIRICHLET (1805–1859) was led to formulate a definition of function, which, if the variable is supposed to represent a number, is substantially the same as that placed at the beginning of this discussion. This definition, which implies that the correspondence may be of any kind whatever, is now generally adopted. Even if we restrict ourselves to the case in which the variable represents a number, this definition does not imply anything regarding an analytic expression for the function in question, or even that such an analytic expression is possible.

Examples of Functions. — Every analytic relation between two variables y and x which associates with every value of x a unique value of y does indeed define a one-valued function as above defined; but such a correspondence may be effected in various other ways. Suppose, for example, that x represents any real number, and let us define the function y to be equal to zero if x represents a rational number and equal to one if x represents an irrational number. This rule defines y as a function of x, according to the definition made. The correspondence may also be brought about by a mechanical device. Let us think, for example, of a thermometer. At any instant of time it indicates a

certain temperature, and this establishes the temperature as a function of the time. If we think of the thermometer as self-recording, this function will be represented graphically by means of a curve from which the number representing the temperature at any instant is readily determined.

The definition we have given does not, however, imply that the classes of the variables be classes of numbers: they may be any classes. If we think of the variable x, for example, as representing any point in the plane, and if we associate with every such point a straight line parallel to a given straight line, we see that these straight lines are functions of the points; or if we consider any two classes whatever which have the same cardinal number, and establish a one-to-one reciprocal correspondence between the elements of the two classes, which is necessary to show that their cardinal numbers are equal, we may say that this correspondence defines a function, in which the variable y representing the elements of one of the classes is a function of the variable x representing the other class.

Other Functions. — We have, for the sake of simplicity, considered only so-called one-valued functions of one variable. A *many-valued* function y of one variable x may be defined as any correspondence whereby to every value of x corresponds a set of values of y. A one-valued *function* z *of two variables* x and y is defined as any correspondence whereby with any value of x and any value of y is associated a single value of z. Here x and y may represent the elements of the same class or of different classes. For example, the pressure exerted by a gas on the sides of an

inclosing vessel is a function of the volume of the gas and the temperature. Or, if we consider the variables x and y as representing points in space, to any value of x and any value of y, distinct from x, will correspond a unique line. The lines of space are thus defined as one-valued functions of pairs of points. It should be noted that we have here an example in which the function is not defined when the two variables represent the same element. The extension of our definition to include the notion of a many-valued function of two or more variables is obvious.

We have already referred to the fact incidentally that a function may in certain cases be represented graphically by means of a curve; and, conversely, it is readily seen that any curve in a plane defines the coördinates of the points on the curve as functions of each other. It was also stated at the outset that this notion of functionality, on account of its fundamental importance not only in mathematics but also in many of the relations of our daily life, should be introduced as early as possible into our mathematics courses. There seems to be no good reason why this conception should not be introduced in a first course in algebra or geometry, if not earlier.[1] The pupil's experience can be drawn on for the most varied examples of this notion, and if judiciously used should contribute much toward arousing his interest in mathematical problems. It goes without saying that the

[1] In fact, it seems desirable to make the notion of a function the central one in a first course in algebra. Cf. E. R. Hedrick, "On the Selection of Topics for Elementary Algebra," *School Science and Mathematics*, vol. XI (1911), p. 51.

graphical method of representing such functions, and geometric methods of defining them should be used wherever possible, on the one hand to emphasize and clarify the notion of function itself, and on the other hand to make clearer and more vivid the relations presented by the problems in question.

A New Cardinal Number. — Before leaving the notion of function, it may be well to make use of it to show the existence of a new cardinal number. It will be recalled that hitherto we have noted the existence of two distinct infinite cardinal numbers, namely, (1) the cardinal number of a denumerable class, and (2) the cardinal number of the continuum; in other words, (1) the cardinal number belonging to the class of all positive integers, and (2) the cardinal number of the class of all ordinary real numbers. We noted that the cardinal number of the latter class was greater than the cardinal number of the former; and it was stated at the time these cardinal numbers were discussed that there existed also cardinal numbers greater than the cardinal number of the continuum. We may now prove this assertion by showing that the cardinal of the class of all possible one-valued functions of a single real variable x is greater than the cardinal number of the continuum. Indeed, suppose it were possible to establish a one-to-one correspondence between the class of all functions $f(x)$ of a variable x and the class of all real numbers R. This would mean that corresponding to any real number r there will exist a unique function of x, say $f_r(x)$. To show that such a correspondence is not possible, we will

actually construct a function which is not contained among the set of functions $f_r(x)$. For this purpose, let us consider the set of functional values obtained by substituting in each of the functions $f_r(x)$ that value of x given by the subscript of the function. For that value of x, each of the functions has a unique value, namely $f_r(r)$.[1] The set of values thus obtained defines a function which we may call the *diagonal function* in this scheme of correspondence, and which we will denote by the symbol $f_x(x)$. Suppose that $F(x)$ is any function of x which differs for every value of x from the function $f_x(x)$ just obtained; for example, let us set $F(x) = f_x(x) + 1$. The function $F(x)$ thus defined is readily seen to be different from every function $f_r(x)$, for it certainly differs from every such function for the particular value $x = r$.[2]

It is interesting to note in passing the analogy between this proof and the proof previously given to show that a continuum cannot be put into one-to-one correspondence with the class of all positive integers (Lecture VIII). It was stated on a previous occasion that the ascending

[1] If $f_r(x)$ does not have a unique finite value, when $x = r$, we may suppose $f_r(r) = 0$ in the argument that follows.

[2] This proves the proposition stated above, that however a correspondence is established between the class of all real numbers and functions taken from the class of all one-valued functions, there will always remain in the latter class functions not affected by this correspondence. In other words, the argument has shown that the cardinal number of the class of all one-valued functions of a real variable is greater than the cardinal number of the continuum. The proof given above is due to G. CANTOR.

sequence of infinite cardinal numbers is unlimited. It is interesting to note further that three of these distinct infinite cardinal numbers can be defined in terms of the simple notions, positive integer, real number, one-valued function of a real variable, which are current in elementary mathematics.

LECTURE XX

LIMIT

The Notion of Limit in Elementary Mathematics. — Closely associated with the notion of variable is the notion of a limit. This notion appears in elementary mathematics both in algebra and in geometry. In the usual first course in algebra it is usually and properly avoided. In a second course in algebra the term occurs incidentally in connection with the study of geometrical progression and in the discussion of irrational numbers. We would venture the opinion, however, that its use in the latter connection is, for pedagogical reasons, undesirable in a high-school course. We find the notion in geometry again, in connection with the consideration of incommensurable ratios, a place where again it should be avoided, and in connection with the discussion of the length and area of a circle, and the surface and volume of a sphere. In view of the fact that a course in geometry usually precedes the second course in algebra, it is probably here that the pupil first comes face to face with the notion of limit. In considering the length of a circle he is asked to consider the inscribed and circumscribed regular polygons and note that the length of the circle is approached more and more closely by the perimeter of the polygons in question as the number of sides of these

polygons is indefinitely increased. He is told that the length of the circle is the limit which the perimeters of the polygons approach when the number of sides is increased indefinitely. This has led, doubtless, to the thoroughly unsatisfactory definition of a limit, which is current in the majority of our textbooks, and which it seems almost impossible to eradicate from the mind of the pupil when once introduced, viz., that "the limit of a variable is a number which the variable may approach infinitely near to, *but which it never reaches.*" It is the italicized portion of this definition which is chiefly objectionable. It is greatly to be desired that our textbook writers and the teachers who use them should have a proper conception of this fundamental notion. It seems desirable for this reason to discuss it in some detail.

A Definition. — A definition of a limit which for most purposes of elementary mathematics will be found satisfactory is the following: If a variable x represents any one of an infinite sequence of numbers $a_1, a_2, a_3, \cdots, a_n, \cdots$, it is said to approach a *limit*, if a number a *exists* such that the numerical value of the difference $x - a$ becomes *and remains* less than any previously assigned positive number ϵ. A more precise formulation of this definition will be given presently, but one or two things may be noted at this point. In the first place, the definition implies, in view of the first italicized phrase, that a variable may not approach any limit; in other words, there may not exist a number a with the specified property. Secondly, it will be observed that the definition implies absolutely nothing as to whether the

variable in question is ever equal to its limit. All that is required is that the numerical value of the difference between the variable x and the limit a shall become and remain less than *any* previously assigned positive number, and nothing implies that this difference may not be zero for some (or indeed all) of the values of x. The phrase "becomes and remains" needs perhaps to be explained a little more fully. The idea implied by the above formulation of the definition is that the variable represents in succession the numbers of the sequence, and the phrase in question then means simply that, *after* a certain point has been reached, depending in general on the number ϵ chosen, the difference between a and every succeeding value of x shall be numerically less than ϵ.

A simple illustration will make this clear. Suppose x represents any number of the sequence

$$\tfrac{1}{2},\ \tfrac{3}{4},\ \tfrac{7}{8},\ \tfrac{15}{16},\ \cdots.$$

The nth term of this sequence is readily seen to be $1-\dfrac{1}{2^n}$, and the numerical value of the difference between the number 1 and the nth term of this sequence is at once seen to be $\dfrac{1}{2^n}$. Now, if *any positive* number ϵ be chosen, however small this number may be, it is easy to see that after n has reached a certain value depending on the value of ϵ, for every value of n greater than this particular value, the difference $\dfrac{1}{2^n}$ will be less than ϵ. The number 1 therefore satisfies the definition of a limit for the sequence considered. This example shows us how to formulate the defini-

tion in question more precisely as follows: If a variable x represents any number of a sequence $a_1, a_2, a_3, \cdots, a_n, \cdots$, it is said to approach a number a as a *limit*, provided that, corresponding to every positive number ϵ, there exists a number m such that the numerical value of the difference $a - a_n$ is less than ϵ, provided only n is greater than m. It seems desirable to give further examples of sequences which satisfy and which do not satisfy this definition. We have just given one which satisfied it. In it the numbers of the sequence were all less than the limit. The sequence

$$\tfrac{3}{2}, \tfrac{5}{4}, \tfrac{9}{8}, \tfrac{17}{16}, \cdots$$

is an example of a sequence which approaches the number 1 as a limit, in which the value of the variable is always greater than the sequence. If we consider the sequence obtained by taking alternately a number from each of the two sequences just considered, namely, the sequence

$$\tfrac{1}{2}, \tfrac{3}{2}, \tfrac{3}{4}, \tfrac{5}{4}, \tfrac{7}{8}, \tfrac{9}{8}, \cdots,$$

we have an example of a sequence which satisfies the definition of having the limit 1, and in which the numbers of the sequence are alternately less than and greater than the limit. This should serve to emphasize the fact that the numbers forming the sequence need not form a sequence in their order of magnitude, as is sometimes supposed. Let us consider now a sequence obtained from the last by inserting the number 1 after every successive pair of terms

$$\tfrac{1}{2}, \tfrac{3}{2}, 1, \tfrac{3}{4}, \tfrac{5}{4}, 1, \tfrac{7}{8}, \tfrac{9}{8}, 1, \cdots.$$

We have here a sequence which still satisfies all the requirements of the definition, of having the limit 1, and in

which for certain values of the variable, namely, every third one, the difference between the variable and the limit is equal to zero; in other words, a sequence in which the variable is sometimes equal to its limit. As a rather trivial, but nevertheless instructive, example of a sequence let us consider one consisting entirely of 1's, namely

$$1, 1, 1, 1, 1, \cdots.$$

This sequence also satisfies the definition of limit, in which the variable is always equal to its limit. The variable is in this case a constant; but we have seen that according to our definition of a variable, a constant is merely a special case of a variable; on the other hand, the sequence

$$1, 2, 3, 4, 5, \cdots$$

clearly does not approach a limit. Furthermore, the sequence

$$1, \ -1, \ 1, \ -1, \ 1, \ -1, \ \cdots$$

does not approach any number as a limit, although its terms are all numerically less than a finite number (*e.g.* 2).

Limits in Geometry.—To return for a moment to the consideration of the length of a circle as the limit of the inscribed or circumscribed regular polygons, it may be noted that, from a logical point of view, the first requirement is to prove that the perimeters of these polygons approach a limit, and then that the length of the circle is *defined* to be this limit. Our geometric intuition seems to furnish us in this case with a proof of the existence of the limit, and for pedagogical reasons this intuitional conception must be regarded as sufficient in a first course

in geometry. It should be noted, however, that our geometric intuition may be a deceptive guide in such matters. That this is so may be made clear from the following example, which will furnish us with an illustration of a limit of a different kind. We formulated our definition above with reference to a variable which represents numbers. Since a variable may represent other things as well, it is natural to inquire as to what meaning may be attached to the term *limit* in such cases. Suppose we have a sequence of curves. Can we formulate a definition of a limit-curve of such a sequence? Let the curves of the sequence be represented by C_1, C_2, C_3, $\cdots$, and suppose there exists a curve C such that the maximum distance between a curve of the sequence and the curve C, measured in some given direction, becomes and remains less than any previously assigned positive number. We may then say that the curve C is the limit of the sequence $C_1, C_2, C_3, \cdots$. As an example, let us consider an equilateral triangle ABC, and let C_1 of our sequence be the broken line ABC (Fig. 25). Bisect the sides of the triangle in the points D, E, F, and let C_2 of the sequence be the broken

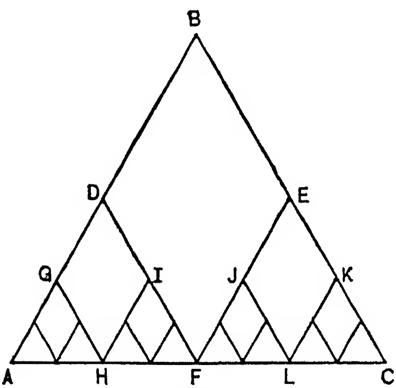

FIG. 25

line $ADFEC$. Bisect the sides of the two new triangles having their bases on the line AC, and let the curve C_3 of our sequence be the broken line $AGHIFJLKC$. Suppose this process continued indefinitely. The maximum distance of any point of one of the broken lines C_1, C_2, C_3, $\cdots$ from the base AC of the triangle, measured perpendicular to this base, is readily seen to become and remain less than any previously assigned positive number ϵ. We may then say that the line AC is the limit of the sequence of broken lines which we have been considering. Does it follow from this, however, that the lengths of the broken lines C_1, C_2, C_3, $\cdots$ have as a limit the length of the line AC? Obviously not; for, if the length of a side of the original equilateral triangle is taken to be unity, it is readily seen that the length of every one of the broken lines C_1, C_2, C_3, $\cdots$ is equal to two. The sequence of lengths is then simply a sequence

$$2, 2, 2, 2, 2, \cdots,$$

the limit of which is the number 2, and is by no means equal to 1, the length of the line AC. And yet our geometrical intuition, if it has not been refined by careful study, might well lead us to assume that the limit of these lengths was indeed equal to the length AC. This example should again emphasize the important fact that there can be no such thing as a "proof" that the limit of the lengths of the inscribed (or circumscribed) polygons mentioned above is the length of the circle, until "length of circle" has been defined. The usual process consists, as has been said, in proving the existence of the limit of the lengths of the poly-

gons and then to *define* this limit to be the length of the circle.

A More General Definition. — The definition of limit which we have given depends essentially on the existence of a sequence of values which the variable represents. In fact, what we have defined is simply the *limit of a sequence.* It is often necessary in mathematics to consider limits of variables which do not represent discrete sequences of the kind previously considered. A more general definition of limit, which will apply to such cases, may be obtained as follows. Let C be any linearly ordered class, and let the variable x represent any element of this class. A *segment* of such a class may be defined as the elements of the class which lie between two given elements of the class. Given an element a, which need not be itself an element of the class C, but which is ordered with reference to C (it may be an element of a linearly ordered class C′ containing C), a *neighborhood* or *vicinity* of a is defined as any segment of the class C such that a lies between two elements of this segment. The element a is then said to be a limit element of the class C, provided every neighborhood of a contains elements of C. That this definition includes as a special case the one previously given is readily seen. We must refrain, however, from a more detailed discussion of this more general definition.

Infinity. — This seems to be the place to consider briefly the notion of infinity as it occurs in elementary mathematics. We have considered the notion hitherto only in connection with that of an infinite cardinal number, and have

seen that these infinite numbers form an increasing sequence. This notion of infinite number, however, hardly occurs explicitly in our elementary courses. The symbol ∞, however, is used and often leads to misconceptions and confusion. What is the meaning that attaches to this symbol and in what sense is its use justifiable? These are questions of considerable importance, and we cannot close this series of lectures without answering them. The symbol is first introduced, probably to represent the division by zero, and the pupil is told that any number divided by zero is equal to infinity, a statement which, of course, seems to him very mysterious. If we will recall the definition of division, viz., that to divide a number a by a number b is to determine a number x such that $bx = a$, we note at once that no such number exists when b is equal to zero; that is, there exists no number x which, when multiplied by zero, gives the product a, if a represents any number different from zero. Moreover, if a does represent the number zero, we see that *any* number x will satisfy this equation.

Now we have noted the general tendency of mathematics to remove as far as possible exceptional cases from its theorems and operations. The whole history of the extension of the number system of algebra illustrates this tendency. The negative numbers were introduced to remove exceptional cases in subtraction, the irrational numbers and the complex numbers, to remove exceptional cases in the extraction of roots. It may well be asked, therefore, is it not in the spirit of mathematics to attempt to remove the

exceptional character of division when the divisor is zero? In some cases this is indeed desirable, and the introduction of a new symbol ∞, or symbols, $+\infty$, $-\infty$, to represent division by zero is sometimes of value; but it must be clearly noted that the introduction of such a symbol is impossible without the violation of some of the fundamental laws characterizing the algebraic symbolism. *Such a symbol cannot, therefore, be regarded as a number in the sense previously defined, and, if introduced at all, must be thought of as satisfying different laws, which form exceptions to the fundamental laws of algebra.* By the introduction of such a symbol with the purpose of removing one exception, numerous other exceptional properties are introduced. The introduction of such a symbol in this sense, therefore, largely defeats its own purpose. There is another meaning, however, which may be attached to this symbol, and which is indeed the meaning in which it is generally used in algebra, though, as we have said, this meaning is often obscured. The use of the symbol ∞ which we have now in mind is closely associated with the notion of limit. If we consider the fraction $\frac{a}{x}$, where we suppose a to represent a number different from zero, and x a variable, we shall have corresponding to every value of x different from zero a unique value of the fraction $\frac{a}{x}$. It is then readily seen that as the variable x approaches the limit zero over any sequence of values not including zero, the fraction $\frac{a}{x}$ increases indefinitely in numerical value. To express this

fact, we may say that $\frac{a}{x}$ "*becomes infinite*" as x approaches zero. *If then we write* $\frac{a}{0} = \infty$, *this expression must be regarded, not as a statement that a divided by zero equals infinity, but merely as a short way of expressing the fact that as the variable x approaches zero, the numerical value of the corresponding variable* $\frac{a}{x}$ *increases indefinitely.* From this point of view, *infinity is a variable.* There is nothing mysterious in such a statement. The introduction may be convenient on account of its brevity, but much care should be employed in its use. In fact, the use of such a symbol in an elementary course seems on pedagogical grounds to be open to serious question. It cannot be too strongly emphasized that *the expression* $\frac{a}{0}$ *is absolutely meaningless,* when thought of as a division by zero, and our pupils should be led from the outset so to regard it.

We had occasion a moment ago, in discussing division by zero, to consider what meaning the definition of division assigns to the notion of dividing zero by zero. We note that this operation involves the determination of a number x such that $x \cdot 0 = 0$. The determination of such a number x is in this case not impossible, but is indeterminate in view of the fact that any number will satisfy this definition. The symbol $\frac{0}{0}$, therefore, may be regarded as a *symbol of indeterminateness.* Its use, however, as expressing an actual division of zero by zero is strictly to be avoided. Such expressions occur in algebra, usually in considering

the quotient $\frac{y}{x}$ of two variables, y and x, each of which approaches zero. It is then possible to inquire whether or not the fraction $\frac{y}{x}$ approaches a limit when y and x each approach zero, and if so, what this limit is. Special examples serve to show that the existence of such a limit and its value depend essentially upon the way in which y and x approach zero. Indeed, the question here raised is meaningless until a correspondence is established between the variables y and x, so that whenever x represents some number, y also represents a definite number; in other words, that the value of the fraction itself is determined as soon as the value of one of the variables y or x is known. y must then be a function of x, or we cannot answer the question raised. To give a few examples, we may note that if y is regarded in succession as being equal to x, or $2x$, or $\frac{1}{2}x$, or x^2, or $\sqrt{x}$, that then as x approaches zero, y will also approach zero, and the fraction $\frac{y}{x}$ will, in each but the last of these cases, approach as a limit respectively the numbers 1, 2, $\frac{1}{2}$, 0, and in the last case becomes infinite.

Another simple and instructive example of this kind is obtained by considering the fraction $\frac{x^2-a^2}{x-a}$. For every value of x differing from a, the denominator will be different from zero, and the fraction will therefore have a determinate value equal to $x+a$. When $x=a$, the denominator and the numerator are both zero, and the operation implied in determining the value of the fraction is indeterminate. If, however, we inquire as to the limit approached by the

fraction as x approaches the limit a, the fraction is readily seen to approach as a limit the number $2a$. In fact, so long as x is different from a, we may write

$$\frac{x^2 - a^2}{x - a} = x + a,$$

and if x is allowed to approach a over any sequence of values (not including a), the quotient may be made to differ from $2a$ by less than any positive number however small. For example, suppose that x approaches a over the sequence of values

$$2a,\ \tfrac{3}{2}a,\ \tfrac{4}{3}a,\ \tfrac{5}{4}a,\ \cdots,\ \frac{n+1}{n}a.$$

The numerator then assumes in turn the sequence of values

$$3a^2,\ \tfrac{5}{4}a^2,\ \tfrac{7}{9}a^2,\ \tfrac{9}{16}a^2,\ \cdots,\ \frac{2n+1}{n^2}a^2;$$

and the denominator in turn the sequence of values

$$a,\ \tfrac{1}{2}a,\ \tfrac{1}{3}a,\ \tfrac{1}{4}a,\ \cdots,\ \frac{1}{n}a.$$

The ratio of the *corresponding* values of the numerator and denominator then assumes in turn the sequence of values

$$3a,\ \tfrac{5}{2}a,\ \tfrac{7}{3}a,\ \tfrac{9}{4}a,\ \cdots,\ \frac{2n+1}{n}a.$$

As n increases indefinitely, the general term of this sequence, which is

$$\frac{2n+1}{n}a = \left(2 + \frac{1}{n}\right)a,$$

approaches the value $2a$.

This process, may, moreover, be given a very vivid geo-

metric interpretation. If we place the fraction in question equal to y, the usual graphic representation of the function

$$y = \frac{x^2 - a^2}{x - a},$$

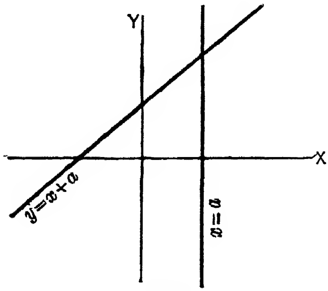

FIG. 26

referred to a system of rectangular cartesian coördinates, is represented, for all values of x differing from a, by the straight line

$$y = x + a.$$

When $x = a$, as we have seen, any value of y will satisfy the equation, and is therefore represented by the points of the line $x = a$. By referring to the graph (Fig. 26), it is readily seen that the value of y is entirely indeterminate when $x = a$, but that the values of y, as x approaches a, do indeed approach the value $2a$.

LECTURE XXI

GENERAL CONCLUSIONS

What is Mathematics? — We have now considered the more important fundamental concepts of elementary mathematics. It seems desirable to devote this last lecture to certain general considerations which could find no place in the previous lectures and which indeed to a large extent presuppose the results which have there been obtained.

We may in particular return to the general questions raised in the first lecture. At the very outset we were confronted with the question: What is mathematics? We should now be in a position to appreciate the difficulties in the way of giving a satisfactory definition It is true that we have considered only the fundamental concepts of elementary mathematics. But we have not considered them from the point of view of elementary mathematics merely. The notion of "number," for example, has been discussed not alone as it occurs in arithmetic and elementary algebra; we have rather considered the meaning that attaches to this term in the most general sense in which it is used throughout the whole of mathematics. We have seen how this meaning has been extended step by step, until "number" appears merely as a symbol representing the elements of a

class which are subject to certain very general laws of combination. A similar state of affairs exists with reference to the majority of the other concepts considered. Our conception of these notions is then sufficiently broad to enable us to consider intelligently the question raised above. But the tendency of mathematics just noted to generalize its concepts to the utmost serviceable limit is precisely one of the elements which render the formulation of a definition of mathematics so difficult.

The Science of Quantity and Space — The old and for a time widely accepted definition to the effect that "mathematics is the science of quantity and space" is utterly inadequate. Even the notion of number, as we have seen, does not necessarily imply anything regarding quantity.

Three Methods. — Three methods suggest themselves for obtaining an answer to the question. As Professor M. Bôcher remarks:[1] "We may seek some hidden resemblance in the various objects of mathematical investigation, and having found an aspect common to them all we may fix on this as the one true object of mathematical study. Or we may abandon the attempt to characterize mathematics by means of its *objects of study*, and seek in its *methods* its distinguishing characteristic. Finally there is the possibility of combining these two points of view." Each of these three methods has been employed in the attempt to formulate a definition of mathematics. To consider these various attempts in detail would carry us too far; moreover, an

[1] M. Bôcher, "The Fundamental Conceptions and Methods of Mathematics," *Bulletin of the American Math. Soc.*, vol. XI (1904), p. 115.

excellent discussion of them is easily accessible to American readers in the article by Professor Bôcher referred to above.

Definitions of Peirce and Russell. — A few conclusions bearing on the problem may, however, be drawn directly from the results of the preceding lectures. In our definition of a "mathematical science" in the first lecture we evidently followed the second of the above methods, since we there regarded the formal logical method as the characteristic feature of a mathematical science. That definition is essentially the same as Benjamin Peirce's, who said that "mathematics is the science that draws necessary conclusions." It is equivalent also to one of Bertrand Russell's definitions to the effect that mathematics is the class of all propositions of the form "P implies Q." We have already criticized these definitions as being too broad; they include more than is ordinarily included in the term *mathematics.* We shall see presently that from another point of view they are too narrow.

Objects of Mathematical Study. — If we adopt the first of the above methods and look for a common element among the objects of mathematical study, we observe that each of the various "branches" of mathematics has for its object the study of a certain class of elements. The elements of these classes are assumed to bear certain relations to each other.[1] If we regard this as the characteristic feature of mathematics, we again get a definition which is too broad. And if we combine the two methods and formulate a defini-

[1] The notion of an "operation" may also be considered as a relation. Thus the operation $aob = c$ is equivalent to a relation between a, b, c.

tion as follows: Mathematics is the science which considers by formal logical methods certain relations assumed to hold between the elements of certain classes, we also find the result too broad; and from another point of view too narrow. In order to render the last definition more satisfactory, we might attempt to find a further resemblance among the classes and among the relations with which mathematics deals with a view to further limiting the content of the definition. But all attempts to discover such a common element in these classes and relations has hitherto failed. The further application of the first method therefore seems to hold out but small prospect of success.

Logical Difficulties.—If then we turn to the second method, that of attempting to define mathematics by means of its methods, we are confronted with two difficulties. If we assume the characteristic method of mathematics to be formal logic, *i.e.* the drawing of necessary conclusions, we must determine with precision what is a *necessary* conclusion. The history of our science shows that the standard of logical rigor has changed with the times. The mathematicians of earlier periods regarded as valid arguments which to-day are not so regarded. Even GAUSS, one of the most critical of the mathematicians who flourished at the beginning of the last century, published proofs which at the present time would not be regarded as rigorous. The standard of rigor at present is doubtless greater than it has ever been. We have seen that it requires that geometric intuition be entirely banished as not sufficiently trustworthy; that indeed an argument to be rigorous must be abstract.

Have we any reason to suppose that we have now reached the limit of logical rigor? If we recall the fundamental logical notions which have been employed, we will note that they are the notions of a "class," of "belonging to a class," and of "correspondence." Are these notions so clear and precise that the reasoning based on them is open to no question? That they are not, the following apparent paradox will show. Attention has recently been called to it by RUSSELL.

A Paradox. — All the classes we have hitherto considered have had the property that no class is an element of itself. Without raising the question as to whether there really exists a class which is an element of itself, *i.e.* such that the class contains itself as an element, let us agree to call any class which does not contain itself as an element an *ordinary class.* Now consider the class C consisting of *all* ordinary classes. Suppose first that C is an ordinary class. It would then follow from the definition of C that C is an element of C; which clearly contradicts the hypothesis that C is an ordinary class. Suppose, then, that C is not an ordinary class. It is then clear that C must be an element of C; for otherwise C would be ordinary. But again from the definition of C, every element of C *is* an ordinary class. We see then that there is a contradiction inherent in the notion of "the class of all ordinary classes," which at first sight seems to be a well-defined class The difficulty seems to be that in defining the class of all ordinary classes we are committing a vicious circle in that we are defining a class in terms of itself. But this does by no

means explain away the fundamental difficulty. How can we determine of any given class whether or not it involves a contradiction? It seems that with reference to the classes ordinarily considered in mathematics all we can say is that, in spite of the enormous amount of work that has been done on them, no contradiction has ever appeared. Our belief in the validity of our reasoning appears to rest on nothing more substantial than this. This is one difficulty in the way of defining mathematics as the science which draws necessary conclusions.

Formal Logic not the Only Method of Mathematics. — There is another. We have seen that a strictly logical treatment of mathematics implies a strictly abstract treatment. From this point of view the objects with which mathematics deals are mere symbols devoid of content except such as is implied in the assumptions concerning them. Do we really maintain that this abstract symbolism constitutes the whole of mathematics? Some one is sure to protest. "What you are considering," he will say, "is surely only the dry bones of the science. It constitutes the whole of mathematics no more than the skeleton of a man constitutes the whole of him." A defender of the abstract point of view might urge that he is defining "pure" mathematics; that everything else which is commonly included in the term is strictly speaking "applied" mathematics, in so far as it deals with concrete applications of an abstract science.

A New Definition of Mathematics. — While much may be said in favor of this attitude, we must admit that in adopting it we are arbitrarily limiting the meaning of the term

"pure mathematics" far beyond the meaning that has attached to it for centuries; and it may seriously be questioned whether it is justifiable or desirable to do so. With this new meaning of the term little if any "pure" mathematics existed fifty years ago. Would it not be better to use the term *abstract mathematics* to denote such an abstract symbolism, or to call it an *abstract mathematical system.* Mathematics as a whole might then be defined as consisting of *all such abstract mathematical systems together with all their concrete applications.*

This definition would seem to include everything that can properly be called mathematics. It should be noted in particular that it includes every concrete science which is capable of abstract formulation. It does not therefore imply that the purely logical method is the only one that may be employed in mathematical investigations. And this would seem to be an essential in any satisfactory definition. For with all our insistence on the formal logical procedure, the important fact must not be lost sight of that formal logic is in only a small minority of cases the method of mathematical discovery. Imagination, geometric intuition, experimentation, analogies sometimes of the vaguest sort, and judicious guessing, these are the instruments continually employed in mathematical research. It is for this reason that we stated above that any definition which sees in the abstract logical method the one characteristic feature of mathematics must of necessity be too narrow. But any proposition, by whatever method it may have been obtained, must be capable of an abstract formulation and a formal

deduction. This appears to be the proper function of the logical method in a definition of mathematics.

We may now restate in more accurate form the definition of a mathematical science given in the first lecture: A mathematical science is any body of propositions which is capable of an abstract formulation and arrangement in such a way that every proposition of the set after a certain one is a formal logical consequence of some or all the preceding propositions. Mathematics consists of all such mathematical sciences. It may still be urged that this definition is too broad. But this objection does not appear to us very serious. Of the actually existing sciences satisfying this definition all will probably justify the term *mathematical* with the current meaning of this term. That it includes, latently as it were, other sciences as yet unborn to which we might not be willing to apply the term *mathematical*, need not concern us until some one really exhibits such a science.

Axioms and Postulates. — Given a mathematical science in its abstract form, we have seen that a certain number of its propositions at the beginning of the sequence are unproved. These are the so-called axioms or postulates of the science. From the strictly logical point of view they are arbitrary assumptions. These assumptions involve certain undefined terms (elements of certain classes and certain relations). These terms from the strictly logical point of view are mere symbols with no other content than is implied in the assumptions concerning them. We have seen further that these assumptions are logically quite

arbitrary, except that they must satisfy the requirement of logical consistency. Here we have the answer to the question raised in the first lecture as to the significance of the axioms and postulates: They are not self-evident truths, neither are they experimental facts. They are assumptions.

Objections to the Abstract Point of View. — Professor KLEIN has recently made a vigorous protest against this point of view. To regard the objects of mathematical study as mere empty symbols sounds the death knell of all science, he says.[1] He recalls the witty though uncomplimentary characterization recently made by Professor THOMAE of men who concern themselves exclusively with meaningless symbols and empty assumptions concerning them. Thomae dubbed such men "thoughtless thinkers." The axioms of mathematics are not arbitrary assumptions, Klein urges; but they are rather *sensible statements* (vernünftige Sätze). He seems to fear that the adoption of the abstract point of view will turn the attention away from the all-important possibility of concrete applications. This fear seems to us groundless. Any one who should devote himself to the development of an abstract symbolism with no reference to its possible concrete applications would indeed deserve the epithet of Thomae. If such a thoughtless thinker really exists anywhere, he has deservedly remained unknown. Professor Klein is doubtless right in urging that the axioms are

[1] F. KLEIN, *Elementarmathematik vom höheren Standpunkte aus*, vol. II, p. 384. See also, with reference to the following discussion, vol. I, pp. 34 ff. of this extremely suggestive work.

sensible statements, not, however, with reference to their logical significance, but because their choice is guided by an intelligent *purpose*.

Abstract Formulation not Sufficient.—Professor Klein has called attention, however, to an important matter which is sometimes overlooked in this connection. We have just recalled the fact that the important question of the logical consistency of a set of assumptions can be determined at present only by exhibiting a concrete application. This is a fatal defect in any attempt to place the foundations of mathematics on a purely formal logical basis. But assuming even that an absolute test of consistency were to be discovered, the problem of exhibiting the logical foundations of a given branch of mathematics, of geometry let us say, would by no means be completely solved by exhibiting a consistent and categorical set of assumptions for the abstract science underlying geometry. We must not lose sight of the fact that the concrete applications of such an abstract system constitute by far the most important part of mathematics; they constitute indeed the "raison d'être" of mathematical study. Having exhibited a logically consistent set of assumptions for what we call abstract geometry, there still remains the essential problem of showing that the intuitive notions of space with which we are familiar do constitute a concrete application of this abstract system. We have here a problem of great difficulty which can certainly not be solved by means of formal logic alone. It is of metaphysical nature; its solution must be sought in the field of philosophy.

The Unity of Mathematics. — Before closing this discussion we would call attention to the fact that the abstract point of view, to the development of which this series of lectures has been devoted, finds what is perhaps its most interesting and important application quite outside the field of the logical foundations proper. We have indeed referred to this fact incidentally in a previous lecture. We may regard the undefined terms in an abstract science as symbols representing *any* entities for which the fundamental assumptions appear to be satisfied. A given abstract science may have many different concrete applications. Several theorems from widely differing branches of mathematics may thus appear merely as different aspects of the same abstract theorem. We must refrain from any detailed discussion. Suffice it to say, that far from being the death of all science, the development of abstract methods during the past few years has given mathematics a new and vital principle which furnishes the most powerful instrument for exhibiting the essential unity of all its branches. Postulational methods have been used from this point of view notably by FRECHET[1] in France and E. H. MOORE[2] in this country with telling effect.

[1] M. FRÉCHET, *Sur quelques points du calcul fonctionnel*, Paris thesis, 1906, reprinted in Rendiconti del Circolo Matematico di Palermo, vol. 22 (1906), pp. 1–74.

[2] E. H. MOORE, *Introduction to a form of general analysis*, The New Haven Colloquium Lectures, Yale University Press (New Haven, 1910).

THE GROWTH OF ALGEBRAIC SYMBOLISM

By U. G. Mitchell

Three Stages in Algebraic Notation. — About seventy years ago Nesselman[1] characterized the historical development of algebraic notation as marked by three stages. (1) *Rhetorical algebra,* in which problems are solved by a course of logical reasoning expressed entirely in words without the use of abbreviations or algebraic symbols. (2) *Syncopated algebra,* in which abbreviations are used for some of the operations and quantities which recur most often. (3) *Symbolic algebra,* in which arbitrary symbols having no apparent connection with the things they represent are employed for all forms and operations. This characterization has proved so helpful and convenient that the terms have gained considerable currency in subsequent writings on the subject.

Rhetorical Algebra. — Although signs for addition, subtraction, and equality are found in the old manuscript of the Egyptian scribe Ahmes[2] (cf. note on p. 101) and

[1] G. H. F. Nesselman, *Die Algebra der Griechen* (Berlin, 1842), pp 302–306.

[2] Cantor, *Vorlesungen ueber Geschichte der Mathematik,* vol. 1 (1894 edition), p. 37.

ARISTOTLE[1] represented quantities by letters in stating problems (but not in equations), it is sufficiently accurate for a general classification to characterize as rhetorical all algebra previous to the time of DIOPHANTUS of Alexandria, who probably lived in the second half of the third or first part of the fourth century of our era.

Among writers of rhetorical algebra whose works appeared after the time of Diophantus may be mentioned the Eastern Arabs, the Western Arabs previous to the thirteenth century, and the early Italian writers including LEONARDO of Pisa (1180–1250?), JORDANUS NEMORARIUS (died in 1236), and their followers as late as the time of REGIOMONTANUS (1436–1476).[2] TROPFKE asserts (vol. I, p. 124) on the authority of WOEPCKE that some of the Arabic writers avoided the use of symbols to the extent of expressing even the numbers in words instead of figures.

Syncopated Algebra. — DIOPHANTUS seems fairly to deserve the credit for initiating syncopated algebra, although it is possible that the notation used by him was known to some of his immediate predecessors and used by them. He wrote a treatise on algebra entitled Αριθμητικα (*i. e.* Arithmetics), in which he made use of signs[3] for subtraction,

[1] Gow, *History of Greek Mathematics* (Camb., 1884), p. 105, note 3.

[2] Cf. TROPFKE, *Geschichte der Elementar-Mathematik* (Leipzig, 1902), vol I, p. 124. See also MATTHIESSEN, *Grundzüge der Antiken und Modernen Algebra* (Leipzig, 1878), p. 269, where a number of examples of notation are given.

[3] According to HEATH, *Diophantos of Alexandria* (Cambridge, 1885), pp. 58–82, the signs were all abbreviations. Gow, in his *History of Greek Mathematics*, p. 109, note, says that he is "inclined to look for the origin of Diophantus' symbols in some hieratic characters"

equality, the unknown quantity, and powers of the unknown quantity as high as the sixth. In the Alexandrian notation current at that time the letters $\alpha, \beta, \gamma, \delta, \epsilon, \ldots, \iota, \kappa$, etc., are used for 1, 2, 3, 4, 5, . . ., 10, 20, etc.

Diophantus indicates addition of numbers by juxtaposition, but a number written immediately after a power of the unknown is the coefficient of that power. For equality he used the sign ἰ and for subtraction a sign ⋔ resembling closely an inverted and truncated ψ. The symbol for the unknown quantity is ς̄ or $ς^{ó}$, and for its second, third, fourth, fifth, and sixth powers the abbreviations are $\delta^{\bar{\upsilon}}$ (for δύναμις, square), $\kappa^{\bar{\upsilon}}$ (for κύβος, cube), $\delta\delta^{\bar{\upsilon}}$ (for δυναμοδύναμις, square-square), $\delta\kappa^{\bar{\upsilon}}$ (for δυναμόκυβος, square-cube), and $\kappa\kappa^{\bar{\upsilon}}$ (for κυβόκυβος, cube-cube), respectively. Diophantus does not go beyond the sixth power of the unknown, and these symbols are not used for the powers of any number other than the unknown. Since addition was indicated by mere juxtaposition, it was necessary in order to avoid confusion that the negative terms should all be placed together after the negative sign and that a sign μ° (abbreviation for μονάδες, units) be written before an absolute term to distinguish it from a variable term. Thus, Diophantus would write

$$\kappa^{\bar{\upsilon}}\bar{\alpha}\ ς^{o\iota}\ \bar{\eta}\ ⋔\ \delta^{\bar{\upsilon}}\ \epsilon\ \mu^{\circ}\bar{\alpha}$$

for $x^3 \cdot 1 + x \cdot 8 - (x^2 \cdot 5 + 1 \cdot 1)$; *i.e.* for $x^3 - 5x^2 + 8x - 1$.[1] It will be noticed that Diophantus' notation was nearly as compact as our own, and in this connection it is worthy of note that in case of mixed numbers (*e.g.* $7\frac{3}{4}$) we still indicate addition by juxtaposition.

[1] Heath, p. 71.

Diophantus represents the highest achievement in algebra of the Greek mind. During the Dark Ages which followed, the greatest advances in algebra were made by the Hindus. In the so-called *Bakshali Arithmetic*, an anonymous manuscript discovered in Bakshali in northwestern India in 1881, and supposed to be a copy of a work written in the third or fourth century, numerical fractions are written in a manner similar to the present practice except that there is no line between numerator and denominator. Integers are written as fractions with denominator one. Addition, division, and equality are indicated by *yu*, *bhâ*, and *pha*, respectively, which are the initial syllables of the corresponding words. Multiplication is indicated by juxtaposition, and subtraction by placing after the subtrahend a cross closely resembling our present plus sign and supposed to be an old form for *ka*, the initial syllable of the word *kanita*, meaning "diminished." Cantor[1] gives the following examples:

$$\boxed{\begin{array}{ccc} 5 & 7 & \\ & & yu \\ 1 & 1 & \end{array}}\ pha\ 12, \text{ for } \tfrac{5}{1} + \tfrac{7}{1} = 12;$$

$$\boxed{\begin{array}{cc} 5 & 32 \\ 8 & 1 \end{array}}\ pha\ 20, \text{ for } \tfrac{5}{8} \times \tfrac{32}{1} = 20;$$

$$\boxed{\begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 3+ & 3+ & 3+ \end{array}}, \text{ for } (1 - \tfrac{1}{3})^3 \text{ or } \tfrac{8}{27}.$$

[1] Vol. I, p. 574.

The unknown quantity is represented by a heavy dot called *sunya*, meaning "empty." This is the same symbol that was used for zero, and their idea seems to have been that the place for the unknown might properly be designated as "empty" until the number to be placed there had been determined.[1]

The later Hindu writers, ARYABHATTA (born 476) and BRAHMAGUPTA (born 598) used the abbreviations *rû* and *yâ* for known and unknown quantities respectively. Addition was indicated by juxtaposition, and in subtraction a dot was placed above the subtrahend. They understood clearly the difference between positive and negative numbers, interpreting them as "possessions" and "debts." Hence the dot used in subtraction is to be taken as a symbol to distinguish negative from positive numbers, rather than as a sign of operation.[2]

Diophantus never used more than one unknown quantity, but the Hindus used several unknowns. They gave them the names of colors and represented them by the initial syllables of the corresponding words. Thus, for the second, third, and fourth unknowns introduced Brahmagupta wrote, respectively, *kâ* (for *kâlaka*, black), *nî* (for *nîlaka*, blue), and *pî* (for *pîtaka*, yellow).

In the English translation of the *Algebra and Arithmetic*

[1] HOERNLE, *Indian Antiquary*, vol. 17, p. 35, quoted by CANTOR, vol. I, p. 574.

[2] CANTOR says: "Das jüngere Pünktchen ist kein Zeichen der Operation, sondern der Zahlenart. Es verwandelt die Subtraktion in eine Addition anders gearteter, entgegengesetzter Grössen." —Vol. I, p. 580.

of Brahmagupta and Bhaskara by H. TH. COLEBROOKE (London, 1817), the following examples[1] of Brahmagupta's notation are given:

$$\begin{array}{lcl} \textit{yá va } 0 \textit{ yá } 10 \textit{ rú } \dot{8} & \text{for} & 0x^2+10x-8 \\ \textit{yá va } 1 \textit{ yá } \;\;0 \textit{ rú } 1 & & =1x^2-\;\;0x+1; \end{array}$$

$$\begin{array}{lcl} \qquad\qquad\qquad\qquad \textit{rú } \dot{9} & \text{for} & \qquad\qquad\quad -9 \\ \textit{yá va } 1 \textit{ yá } 10 & & =1x^2-10x. \end{array}$$

From these examples it will be seen that equality was expressed not by a special sign, but merely by writing the first member of the equation above the second.

To the Hindus belongs the credit of recognizing the existence of irrational numbers and of performing operations with them as with other numbers. Their sign for square root was the abbreviation *ka*, for *karana*, meaning "irrational." BHASKARA (born 1114) has already been quoted (*ante*, p. 107) in regard to operations with negative numbers.

The rise of Saracen learning in the later Middle Ages contributed little to the growth of algebraic symbolism since the Eastern Arabs wrote only rhetorical algebra and the Western Arabs made no use of syncopated algebra before the thirteenth century. In the fourteenth and fifteenth centuries the Western Arabs seem to have developed a considerable symbolism, as may be seen from the work of an Andalusian writer, ALKALSADI, who lived in

[1] Reproduced by CANTOR, vol. I, p. 582, and by TROPFKE, vol. I, p. 130.

the second half of the fifteenth century. The title of his book was *Raising the Veil of the Science of Gubar*, the word *Gubar* being used here figuratively to represent written as distinguished from mental arithmetic. Alkalsadi indicated equality and square root by the final and initial letters respectively of the corresponding words.[1] TROPFKE remarks[2] that by employing an equality sign and a way similar to our own for writing a proportion, he surpassed in some respects the Hindus.

We may take as representative men whose works aided the development of syncopated algebra in Western Europe, NICOLAS CHUQUET in France, JOHANN WIDMAN in Germany, and LUCA PACIOLI in Italy.

Chuquet's treatise, *Le Triparty en la science des nombres*, was written at Lyons in 1494, but never appeared in print until 1880. Nevertheless, according to Cantor[3] it was widely circulated by means of copies and its influence was further extended by the publication at Lyons in 1520 of an arithmetic, a large part of which was taken bodily from Chuquet's manuscript. This arithmetic was written by ESTIENNE DE LA ROCHE, one of Chuquet's pupils, and was the best of early French arithmetics. To Chuquet's *Le Triparty* can be traced the use of our present radical sign with indices to show what root is to be taken. Taking the letter R as an abbreviation for the word *radix*, meaning root,

[1] CANTOR, vol. I, p. 764. Cf. also CAJORI, *History of Elementary Mathematics* (New York, 1910), pp. 110-111.

[2] Vol. I, p. 130.

[3] Vol. II (1892 edition), p. 318.

Chuquet wrote[1] $R^2 16 = 4$, $R^3 64 = 4$, $R^4 16 = 2$, $R^5 243 = 3$. The following examples will serve to illustrate further his notation:[2]

R^2. 13. $\tilde{m}$. R^2. 7. $\tilde{p}$. R^2. 6. par R^2. 5. $\tilde{m}$. R^2. 2., for $\dfrac{\sqrt{13} - \sqrt{7} + \sqrt{6}}{\sqrt{5} - \sqrt{2}}$;

12. plus. 3^2 égaux à. 30^1., for $12 + 3\,x^2 = 30\,x$.

It is particularly worthy of note that he wrote (as illustrated in the second example above) small numbers above and at the right of the coefficient to indicate the first few powers of the unknown. Thus he writes

$$12^0,\ 12^1,\ 12^2,\ 12^5, \text{ for } 12,\ 12\,x,\ 12\,x^2,\ 12\,x^5,$$

and even[3]

$$8^3 \text{ multiplie par } 7^{1m} \text{ monte } 56^2 \text{ for } 8\,x^3 \cdot 7\,x^{-1} = 56\,x^2.$$

This is the first appearance of the zero exponent. Fractional exponents and even some rules for exponents had appeared in an earlier work entitled *Algorismus Proportionum* by a French bishop, NICOLE ORESME, who was born about 1323 and died in 1382. In Oresme's notation $4^{\frac{3}{2}}$ is expressed[4] by

$$\boxed{1\ \mathrm{p}\tfrac{1}{2}}\ 4 \quad \text{or by} \quad \boxed{\frac{\mathrm{p}.1}{1.2}}\ 4.$$

Chuquet may have been influenced by the work of Oresme, but Cantor[5] is of the opinion that such is not the case.

The first great German text-book on arithmetic is that of

[1] CANTOR, vol. II, p. 324.
[2] TROPFKE, vol. I, p. 315.
[3] CANTOR, vol. II, p. 324.
[4] TROPFKE, vol. I, p. 200.
[5] Vol. II, p. 327.

JOHANN WIDMAN, published at Leipzig in 1489, in which the signs + and − appeared for the first time in print. Widman used them to indicate excess and deficiency. For example, if a case of goods was expected to weigh 4 centners and weighed 5 lbs. more or 5 lbs. less, Widman gives its weight as 4c + 5 lbs. or 4c − 5 lbs., respectively.[1] Through the writings of GRAMMATEUS (1518), RUDOLFF (1525), STIFEL (1544), and others the + and − signs came into general use in Germany long before they were adopted elsewhere.

The introduction of syncopated algebra into general use outside of Germany was due, in large measure, to an Italian friar, LUCA PACIOLI, who is said to have lectured on mathematics at Rome, Pisa, Venice, and Milan. His *Summa de Arithmetica, Geometrica, Proporzioni e Proporzionalita* was printed at Venice in 1494, and is important because it was the first great general work on mathematics printed, and attained a wide circulation. In it addition and subtraction are indicated by the initial letters *p* and *m*, respectively, and abbreviations are used for the first 29 powers of the unknown quantity. The absolute term and x, x^2, x^3, x^4, etc., are respectively represented by *numero* or n^o, *cosa* (Ital. for *thing*) or *co*, *censo* or *ce*, *cubo* or *cu*, *censo de censo* or *ce. ce*, etc.[2] The notation for indicating roots is like Chuquet's, except that square, cube, and fourth roots are sometimes rep-

[1] BALL, *A Short History of Mathematics*, fourth edition (London, 1908), p. 207

[2] FINK, *A Brief History of Mathematics*, translation by BEMAN and SMITH (Chicago, 1903), p. 96.

resented by ℞., ℞ *cu.*, and ℞℞, respectively. The quantity $\sqrt{40-\sqrt{320}}$ is expressed by ℞ V 40 m ℞ 320, the ℞ V being an abbreviation for *radix universalis* (general root). BALL says (p. 211) that *ae* (for *aequalis*, equal) is sometimes used as an equality sign, and in two instances in writing a proportion a known number is represented by a letter.

As a later example of syncopated algebra we may notice the *Ars Magna de Rebus Algebraicis* of JEROME CARDAN, published at Nuremburg in 1545, and famous chiefly because it contained the solutions of the general cubic and biquadratic equations. As illustrations of the notation used in the *Ars Magna* we have

cubus *p* 6. rebus aequalis 20,

for $$x^3+6x=20,$$

and R. *u.* *cu.* R. 108 $\bar{p}$. 10 $\bar{m}$. R. *u.* *cu.* R. 108 $\bar{m}$. 10,

for $$\sqrt[3]{\sqrt{108}+10}-\sqrt[3]{\sqrt{108}-10},$$

the value of x as found by solution of the given cubic.[1]

The invention of printing (about 1450) contributed greatly to the development of algebra. Printing gave to the writer of a book a much larger circle of readers, and the wide dissemination of printed translations of Greek, Hindu, and Arabian mathematical works and of books based upon them gave a great impetus to mathematical thought. The second half of the sixteenth century saw symbolic algebra well begun, and by the close of the seventeenth century much of our present symbolism was fully established.

[1] MATTHIESSEN, pp. 364 and 368.

The Beginnings of Symbolic Algebra. — In 1557 ROBERT RECORDE published in London a book, largely algebraic in character, entitled *The Whetstone of Wit*. In this book is found for the first time in print the sign of equality (=). In speaking of the sign Recorde says: "And to avoide the tediouse repitition of these woordes: is equalle to: I will sette as I doe often in woorke use, a paire of parallelles, or Gemowe lines of one lengthe, thus: ═════, bicause noe. 2. thynges can be moare equalle."[1]

The sign ꝏ or ∞, a contraction for the *ae* in *aequalis*, was largely used (notably by DESCARTES) during the seventeenth century, and it was fully a hundred years after its appearance that Recorde's sign came into general use.

Not long after the time of Recorde, BOMBELLI (1572) introduced aggregation signs similar to the brackets now in use, and the Hollander STEVIN (1585) made use of $\sqrt{③}$, $\sqrt{④}$, etc., to indicate cube root, fourth root, etc. Stevin thus added the use of indices to that of a radical sign which had previously appeared in the works of RIESE (1524), RUDOLFF (1525), and STIFEL (1544). Before the close of the century, however, VIETA had published (1591) his algebra *In artem analyticam isagoge* and laid the foundation of modern symbolic algebra. He made the great advancement of representing known as well as unknown quantities by letters. "From that day," says LAISANT, "when the search for values gave way to the search for the

[1] See SMITH, *Rara Arithmetica* (Boston, 1908), p. 288, where is given a photographic reprint of the page on which the quotation occurs.

operations to be performed, the idea of the mathematical function enters into the science, and this is the source of its subsequent progress." [1]

Vieta used consonants for known and vowels for unknown quantities. Thus, he wrote A, Aq, Ac, Aqq, etc., for x, x^2, x^3, x^4, etc., and $+$ and $-$ for addition and subtraction. For $\frac{ax}{b}+\frac{ax-ac}{d}=a$, Vieta put $\frac{B \text{ in } A}{D}+\left\{\frac{\begin{matrix}B \text{ in } A\\ -B \text{ in } H\end{matrix}}{F}\right\}$ aequabantur B, and for $d(2b^3-d^3)$ he put D in $\left[\begin{matrix}B \text{ cubum } 2\\ -D \text{ cubo}\end{matrix}\right.$ [2].

These examples illustrate the fact that he used the bracket and introduced the braces. Vieta's influence was greatly enhanced by his wealth. He had treatises and books printed at his own expense and sent to friends in all lands who were interested in the subject. He is said to have spent "20,000 Thaler in klingender Münze" (*i.e.*, "20,000 dollars in hard cash") in this way.[3]

Following the work of Vieta, advances in algebraic symbolism came rapidly. In 1629, ALBERT GIRARD, a Flemish mathematician, introduced parentheses and the signs $\sqrt[2]{\ }$, $\sqrt[3]{\ }$ for square and cube root. Two years later THOMAS HARRIOT published in London his *Artis Analyticæ Praxis*, in which were introduced the signs $<$ and $>$ with their present meanings. Like Vieta, he used consonant letters for known and vowels for unknown quantities, replacing, however,

[1] LAISANT, *La Mathématique*, p. 55, quoted by SMITH, *The Teaching of Elementary Mathematics* (New York, 1904), p. 156.

[2] TROPFKE, vol. I, p. 327. [3] CANTOR, vol. II, p. 537.

Vieta's capitals with small letters In this work of Harriot's the cross ($\times$) for multiplication appeared almost simultaneously with its introduction by WILLIAM OUGHTRED in his *Clavis Mathematica,* published in the same year (1631). The sign is due to Oughtred. Harriot also made the important improvement of writing aa for a^2, aaa for a^3, etc. For example, he wrote[1]

$$aaa - 3{\cdot}bba = +2{\cdot}ccc$$
$$\text{for } x^3 - 3\,b^2x = 2\,c^3,$$

and $\overline{\sqrt{ccc + \sqrt{cccccc - bbbbbb}}} + \overline{\sqrt{ccc - \sqrt{cccccc - bbbbbb}}} = a,$

for $$\sqrt[3]{c^3 + \sqrt{c^6 - b^6}} + \sqrt[3]{c^3 - \sqrt{c^6 - b^6}} = x.$$

In 1634, three years after the appearance of Harriot's work, PIERRE HERIGONE brought out at Paris a *Cours Mathématique* in 5 or 6 volumes in which a^2, a^3, a^4, etc., were written as $a2$, $a3$, $a4$, etc. It only remained for DESCARTES, in his *Géométrie* of 1637, to put this notation into the form aa, a^3, a^4, etc., and for NEWTON (about 1666) to make the generalization for rational exponents.

Two other improvements in notation are due to DESCARTES — the use of x, y, z for unknown and a, b, c for known quantities, and the introduction of the vinculum and its combined use with the radical sign. This last step may have been suggested by the practice of Harriot as illustrated in the second example given above.

For $\sqrt[3]{\frac{1}{2}q + \sqrt{\frac{1}{4}q^2 + \frac{1}{27}p^3}}$ Descartes wrote[2]

$$\sqrt{C.+\tfrac{1}{2}q + \sqrt{\tfrac{1}{4}qq + \tfrac{1}{27}p^3}}.$$

[1] TROPFKE, vol. I, p. 330. [2] *Ibid.*, p. 140.

The form $\frac{a}{b}$ for common fractions dates back to the Arabs and possibly to the Hindus. Its first use for algebraic expressions dates from a German manuscript of about 1460, where such expressions as $\frac{100}{1 \text{ ding}}$ and $\frac{12 \text{ res et } 45}{1 \text{ census et } 3 \text{ res}}$ are used[1] for $\frac{100}{x}$ and $\frac{12x+45}{x^2+3x}$.

The sign $\div$ is first found in the *Teutsche Algebra* of J. H. Rahn (Zurich, 1659). Leibnitz in 1684 made use of the double dot ( ) for ratio, and to him should be credited also the use of subscripts (1676), and of a single letter to represent a function, as the ξ function of x. Although Leibnitz conceived the idea of determinants, its modern notation belongs to a much later date, chiefly due to Cauchy, about 1812.

Although one readily calls to mind other familiar symbols of algebra (*e.g.*, the sign ∞ for infinity, due to Wallis, 1655, the symbol $n!$ for $1 \cdot 2 \cdot 3 \; . \; . \; . \; \cdot n$, due to Kramp, 1808, etc.), enough has been said to show that by the close of the seventeenth century a sufficient symbolism had been developed to determine the general form of our present usage and to give the subject a language of its own emancipated from subjection to the rules of syntax. From the notation of Leibnitz and Newton to the employment of arbitrary symbols for any form or operation whatever, or even, as in recent times (cf. *ante*, p. 88), for a relation entirely undefined, seems to us an easy step when compared with the difficulties which faced the early explorers in the field.

[1] Tropfke, vol. I, p. 137.

INDEX

Printed in the United States of America.